Propriétés structurelles et calculatoires des pavages

Emmanuel Jeandel

Habilitation à diriger des recherches

présentée devant

L'UNIVERSITÉ MONTPELLIER II

École Doctorale I2S

Propriétés structurelles et calculatoires des pavages

Emmanuel JEANDEL

Présentée le 13 décembre 2011

après avis des rapporteurs :

Olivier BOURNEZ	Professeur à l'Ecole Polytechnique
Jarkko KARI	Professeur à l'Université de Turku
Joseph S. MILLER	Professeur à l'Université de Wisconsin-Madison

devant le jury composé de :

Olivier BOURNEZ	Professeur à l'Ecole Polytechnique
Bruno DURAND	Professeur à l'Université de Montpellier 2
Jarkko KARI	Professeur à l'Université de Turku
Jean MAIRESSE	Directeur de Recherches au CNRS
Joseph S. MILLER	Professeur à l'Université de Wisconsin-Madison
Gwénaël RICHOMME	Professeur à l'Université Paul-Valéry Montpellier 3

NOTATIONS

$\llbracket n \rrbracket$	Entiers de 1 à n (1)
$\mathsf{M} \sqsubseteq \mathsf{N}$	Le motif M est inclus dans le motif N (2)
$C^{\mathbb{Z}^2}$	Ensemble des configurations sur les couleurs C (3)
$C^{[\mathbb{Z}^2]}$	Ensemble des motifs finis sur les couleurs C (3)
$\mathcal{L}(\mathsf{X})$	Ensemble des motifs finis que contient X (3)
$\mathcal{L}_I(\mathsf{X})$	Ensemble des motifs de support I que contient X (3)
$p_I(\mathsf{X})$	Nombre de motifs de support I que contient X (3)
$\mathcal{F}(U)$	Ensemble des motifs n'apparaissant pas dans U (4)
$\mathcal{S}(V)$	Espace de pavages défini par V (4)
$\mathsf{X} \preceq \mathsf{Y}$	Tout motif de X est présent dans Y (16)
$S(T)$	Ensemble des configurations vérifiant T (43)

Table des matières

CV

né le 10 janvier 1980 à Epinal (Vosges), nationalité française, célibataire.

Fonctions occupées

2011–2012	Délégation CNRS au Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM)
2006–	Maître de conférences en Informatique, Université de Provence. Recherche au Laboratoire d'Informatique Fondamentale de Marseille (LIF)
2006	Postdoctorant à l'Institut pour le Calcul Scientifique de Paderborn (PaSCo, Allemagne)
2003–2005	Allocataire de recherche au sein de l'équipe MC2 au LIP, ENS Lyon. Moniteur à l'ENS Lyon.
1999–2003	Elève Normalien de l'ENS Lyon.

Formation

2005	Doctorat d'Informatique de l'École Normale Supérieure de Lyon. Titre de la thèse : *Techniques algébriques en calcul quantique*.
2002	DEA d'Informatique Fondamentale de l'École Normale Supérieure de Lyon (rang : premier).
1997	Baccalauréat série S, mention Bien.

Activités

Responsabilités pédagogiques

2008–2011	coresponsable pédagogique avec Séverine Fratani du M1 Informatique, Université de Provence. `http://www.lif.univ-mrs.fr/~sfratani/master/`
2007–2008	coresponsable pédagogique avec Bruno Durand du L3 Informatique en télé-enseignement, Université de Provence `http://ctes.univ-provence.fr/`

Activités de représentation

2007–	Membre du conseil d'UFR et de la commission enseignement de l'UFR MIM.
2006–2011	Représentant du LIF au conseil de la Bibliothèque de l'UFR MIM.
2003–2005	Représentant des doctorants au conseil de la Bibliothèque, ENS Lyon.
2002–2005	Représentant des doctorants à l'Ecole Doctorale MathIf, Lyon.
2002–2003	Représentant des élèves au conseil de la Bibliothèque, ENS Lyon.

Université

2009–2011	Rapporteur pour les dossiers de promotions hors classe du secteur Sciences.

Equipe

2007–2011	Représentant de l'Equipe Escape pour la mission Valorisation du LIF.
2002–2005	Responsable Informatique (matériel et page web) de l'équipe MC2 (LIP, ENS Lyon).

Collaborations

2009–	Membre de l'ANR blanche EMC.
2010–2011	Membre de l'Action de Recherche Coopérative $CaCO_3$

Collaborateurs nationaux : G. Theyssier (LAMA), N. Ollinger (LIFO), Laurent Bienvenu (LIAFA), Francis Oger (Logique). Collaborateurs Internationaux : R. Pavlov (U. Denver), M. Hochman (U. Princeton), D. Cenzer (U. Floride), J. Kari (U Turku).

Animation de la recherche

Relecture d'articles pour des conférences (ICALP, LICS, STACS, CCC, FSTTCS, JAC, LATIN, MCU, MFCS, SOFSEM, SPAA, ...) et journaux (Theoretical Computer Science, Information and Computation, Information Sciences, Quantum Information and Computation, Discrete Mathematics and Theoretical Computer Science...).

2009–2010	Coorganisateur du mois thématique Mathématiques et Informatique 2010, CIRM `http://www.lirmm.fr/arith/wiki/MathInfo2010/MathInfo2010`
2008	Coorganisateur de la conférence JAC'08 `http://jac.lif.univ-mrs.fr/`
2008	Expertise Nationale dans le cadre de l'Agence Nationale de la Recherche (ANR)

Comités de sélection

2009-2010	MCF 1109 et MCF 1196 *Informatique Fondamentale* (Université de Provence, Laboratoire d'Informatique Fondamentale de Marseille)
2009-2010	MCF 328 *Informatique fondamentale : modèles discrets, systèmes complexes* (Chaire CNRS, Université Nice Sophia Antipolis, Laboratoire d'Informatique Signaux et Systèmes de Sophia-Antipolis I3S)
2009-2010	MCF 0644 *Traitement automatique des langues, Algorithmique* (Université Montpellier 2, IUT Montpellier, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier LIRMM)
2008-2009	MCF 0572 *Logique et interactions* (Université de la Méditerranée, Institut de Mathématiques de Luminy IML, section 25)
2007-2008	MCF 0538 *Mathématiques et Interfaces avec l'Informatique* (Université de la Méditerranée, Institut de Mathématiques de Luminy IML, section 25)

Enseignement

On pourra trouver quelques notes de cours à l'adresse suivante :
`http://www.lif.univ-mrs.fr/~ejeandel/enseignement.html`

2006–	Enseignement statutaire, Université de Provence.
2009–2011	Examinateur à l'épreuve d'entrée du concours des ENS
2010–2011	Enseignement (vacations) à l'Université de Nice-Sophia Antipolis.
2007–2009	Enseignement (vacations) à l'Université de Nice-Sophia Antipolis.
2002–2005	Enseignement (monitorat) à l'Université Lyon I, INSA Lyon et ENS Lyon.

Publications

Journaux

[A1] E. Jeandel. *The periodic domino problem revisited*. Theoretical Computer Science, tome 411, pages 4010–4016, 2010. doi:10.1016/j.tcs.2010.08.017

[A2] P. Charbit, E. Jeandel, P. Koiran, S. Perifel et S. Thomassé. *Finding a Vector Orthogonal to Roughly Half a Collection of Vectors*. Journal of Complexity, tome 24(1), pages 39–53, 2008. doi:10.1016/j.jco.2006.09.005

[A3] E. Jeandel et N. Ollinger. *Playing with Conway's Problem*. Theoretical Computer Science, tome 409, pages 557–564, 2008. doi:10.1016/j.tcs.2008.09.026

[A4] E. Jeandel. *Topological Automata*. Theory of Computing Systems, tome 40(4), pages 397–407, 2007. doi:10.1007/s00224-006-1314-y

[A5] V. D. Blondel, E. Jeandel, P. Koiran et N. Portier. *Decidable and Undecidable Problems about Quantum Automata*. SIAM Journal on Computing, tome 34(6), pages 1464–1473, 2005. doi:10.1137/S0097539703425861

[A6] H. Derksen, E. Jeandel et P. Koiran. *Quantum automata and algebraic groups*. Journal of Symbolic Computation, tome 39(3–4), pages 357–371, 2005. doi:10.1016/j.jsc.2004.11.008

Conférences

[I1] E. Jeandel et P. Vanier. *Π^0_1 sets and tilings*. Dans *Theory and Applications of Models of Computation (TAMC)*, *Lecture Notes in Computer Science*, tome 6648, pages 230–239. 2011. doi:10.1007/978-3-642-20877-5_24

[I2] A. Ballier, B. Durand et E. Jeandel. *Tilings robust to errors*. Dans *Latin American Theoretical Informatics Symposium (LATIN)*, *Lecture Notes in Computer Science*, tome 6034, pages 480–491. Springer, 2010. doi:10.1007/978-3-642-12200-2_42

[I3] A. Ballier et E. Jeandel. *Computing (or not) quasi-periodicity functions of tilings*. Dans *Symposium on Cellular Automata (JAC)*, pages 54–64. 2010. hal.archives-ouvertes.fr/hal-00542498

[I4] E. Jeandel et P. Vanier. *Periodicity in Tilings*. Dans *Developments in Language Theory (DLT)*, *Lecture Notes in Computer Science*, tome 6224, pages 243–254. Springer, 2010. doi:10.1007/978-3-642-14455-4_23

[I5] E. Jeandel et P. Vanier. *Slopes of tilings*. Dans *Symposium on Cellular Automata (JAC)*, pages 145–155. 2010. hal.archives-ouvertes.fr/hal-00542000

[I6] E. Jeandel et G. Theyssier. *Subshifts, Languages and Logic*. Dans *Developments in Language Theory (DLT)*, *Lecture Notes in Computer Science*, tome 5583, pages 288–299. Springer, 2009. doi:10.1007/978-3-642-02737-6_23

[I7] A. Ballier, B. Durand et E. Jeandel. *Structural Aspects of Tilings*. Dans *Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 61–72. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2008. doi:10.4230/LIPIcs.STACS.2008.1334

[I8] A. Ballier et E. Jeandel. *Tilings and Model Theory*. Dans *Symposium on Cellular Automata Journées Automates Cellulaires (JAC)*, pages 29–39. MCCME Publishing House, Moscow, 2008. hal.archives-ouvertes.fr/hal-00273698_v1

[I9] E. Jeandel. *Topological Automata*. Dans *Symposium on Theoretical Aspects of Computer Science (STACS)*, *Lecture Notes in Computer Science*, tome 3404, pages 389–398. Springer, 2005. doi:10.1007/978-3-540-31856-9_32

[I10] E. Jeandel. *Universality in Quantum Computation*. Dans *International Colloquium on Automata, Languages and Programming (ICALP)*, *Lecture Notes in Computer Science*, tome 3142, pages 793–804. Springer, 2004. doi:10.1007/978-3-540-27836-8_67

Articles soumis

[U1] E. Jeandel et G. Theyssier. *Subshifts as Models for MSO Logic*. Soumis à Information and Computation.

[U2] E. Jeandel et P. Vanier. *Turing degrees of multidimensional SFTs*. Soumis à Theoretical Computer Science.

Recherche

Mon activité de recherches est consacrée en majorité à l'étude des modèles de calcul et de leur dynamique. Pendant ma thèse effectuée à Lyon de 2002 à 2005, le modèle étudié est celui des automates et des circuits quantiques. Après mon recrutement dans l'équipe Escape du laboratoire d'Informatique Fondamentale de Marseille, j'étudie principalement un autre modèle de calcul, les pavages. Cependant, bien que les modèle soient différents, mon approche est similaire dans les deux cas. D'abord un point de vue très *algébrique*, où je cherche à étudier ces objets à travers les structures algébriques associées : groupe de Lie pour le calcul quantique, topologie et préordres pour les pavages. Ensuite un point de vue *calculatoire*, visant à estimer très précisément leur puissance en tant que modèles de calcul, à travers des résultats allant si possible au delà de la simple indécidabilité.

Pour plus de lisibilité, les travaux décrits ci-dessous concernent principalement mes travaux actuels sur les pavages, objet de l'habilitation et ne décrivent pas les travaux de recherche effectués antérieurement pendant ma thèse et mon postdoctorat.

Activités de recherche

Le modèle des pavages comme je l'étudie a été introduit par Wang [Wan61] en 1961 pour étudier la satisfiabilité de formules d'un certain fragment de la logique du premier ordre. Le modèle est défini de façon très simple et consiste à essayer de paver le plan discret $\mathbb{Z}^2$ par un nombre fini de tuiles colorées, en respectant des contraintes d'agencement locales.

Bien qu'il s'agisse d'un modèle défini géométriquement, il s'avère assez vite fondamental dans l'étude des modèles de calcul. Les premiers résultats montrent en effet qu'on peut y encoder du calcul [KMW62] et il s'impose comme une alternative viable aux machines de Turing, en particulier pour montrer des résultats d'indécidabilité [vEB97]. Il intervient de plus naturellement lorsqu'on étudie les systèmes complexes, et en particulier les automates cellulaires, pour lesquels les pavages sont un des outils essentiels d'analyse [Kar92, Dur94], puisqu'on les rencontre dès qu'on étudie les diagrammes espace-temps, ensembles limite et points fixes d'automates cellulaires. C'est en particulier dans ce cadre que j'ai commencé à étudier les pavages dans le cadre de l'équipe Escape au LIF et de l'ANR Sycomore.

Lorsqu'on s'intéresse à ce même objet d'un point de vue dynamique, les pavages du plan discret sont également appelés *sous-shifts* (décalages) de type fini, objet d'étude de l'important domaine de la dynamique symbolique [LM95, MH38]. L'apport des pavages du plan discret (et de la droite discrète) est alors essentiel : Ainsi on peut démontrer que tout système dynamique raisonnable (effectif) peut se voir comme l'action du décalage (la translation) sur un espace de pavages, ou encore comme l'action d'un automate cellulaire sur son ensemble limite [Hoc09b, Théorème 1.6]. L'étude des ces objets en dimension 1 est bien comprise, mais il y reste un important problème ouvert, qui est de savoir décider quand deux systèmes dynamiques symboliques sont "isomorphes". Le terme exact est conjugaison, et nous reviendrons sur cette question plus loin.

Dès mon arrivée dans l'équipe Escape, j'ai étudié les pavages suivant ces deux axes, récursif et dynamique.

Structure des pavages

Afin d'étudier la dynamique des pavages, la première approche que j'ai développée est l'utilisation de structures algébriques particulières. J'ai encadré dès mon arrivée au LIF la thèse d'Alexis Ballier, qui a étudié ces questions. Nous avons en particulier développé l'usage d'une relation d'ordre partiel sur les motifs [I7], qui a permis d'obtenir des résultats très précis sur le ensembles de pavages dénombrables. Il s'agissait des premiers résultats dans ce contexte. J'ai ensuite encadré en 2008 un étudiant de L3 (B. Le Gloannec) pour prolonger ce sujet.

J'ai également exploré une autre approche structurelle avec A. Ballier, puisque nous avons étudié dans [I8] les pavages comme modèles d'une théorie universelle. Nous avons à cette occasion collaboré avec F. Ogier (Equipe de Logique Mathématique), auteur d'un article similaire [Oge04]. Cela l'a conduit à montrer comment notre approche par ordre partiel pouvait être incorporée à l'approche théorie des modèles dans un article [Oge] présenté au Logic Colloquium 2010.

Ces travaux sur la logique ont donné lieu à un cours à l'École Jeunes Chercheurs en Informatique Mathématique à Marseille en 2008, puis à diverses présentations à des groupes de travail. A cette occasion, j'ai collaboré avec G. Theyssier (LAMA) et nous avons montré comment obtenir une caractérisation type logique descriptive des espaces de pavages sofiques, résultat présenté à DLT [I6] et actuellement prolongé [U1].

J'ai enfin organisé en février 2010 avec N. Bedaride, J. Cassaigne, Th. Fernique, Th. Monteil et M. Sablik un mois thématique sur les interactions entre mathématiques et informatique. Nous avons organisé le mois mais également proposé les différents orateurs et thèmes de cette manifestation. Deux des cinq semaines étaient en lien direct avec les thèmes présentés ici, plus spécifiquement une semaine était consacrée aux liens entre dynamique et calcul, et une semaine consacrée aux pavages. A cette occasion, j'ai donc pu communiquer avec différents spécialistes internationaux du domaine. J'ai ainsi donné un contre-exemple (non publié) montrant l'optimalité d'un résultat de Michael Hochman (Université de Princeton/Université Hébraique de Jérusalem). J'ai présenté une question ouverte sur laquelle j'ai travaillé avec Ronnie Pavlov (Université de Denver), qui a obtenu quelques résultats préliminaires [Pav11].

Récursivité

Dès mon arrivée au LIF, j'ai participé à la création d'une ANR Blanche, sur l'émergence dans les modèles de calcul (EMC). Cette ANR est rassemblée autour de l'équipe Escape du LIRMM/LIF et du pôle MDSC de l'I3S, ainsi que d'autres chercheurs associés (GREYC, LIFO, LAMA). Je suis plus particulièrement chargé d'une tâche sur les liens entre dimension et dynamique. Afin d'étudier le problème de la conjugaison (savoir si des jeux de tuiles donnent des ensembles de pavages "isomorphes"), les dynamiciens ont introduit de nombreux invariants de conjugaison (quantités égales pour des jeux de tuiles conjugués). Ces invariants sont bien compris en dimension 1, et un des buts de la tâche est de les caractériser en dimension supérieure.

J'ai pour cela encadré le stage de Master de Pascal Vanier en 2009, où nous avons montré [I4] comment l'ensemble des périodes possibles (un invariant de conjugaison) est relié à la question ($\mathbf{P} \stackrel{?}{=} \mathbf{NP}$). J'ai ensuite décidé d'encadrer la thèse de Pascal, sur le sujet précis des invariants de conjugaison. Nous avons ainsi dans [I5] caractériser un autre invariant, les directions de périodicité.

Dans le cadre d'une autre tâche de l'ANR EMC, consacrée aux petites machines, j'ai également encadré le stage de L3 de N. Rolin en 2010, où on a montré l'indécidabilité de la pavabilité du plan par un nombre fini de barres de Wang (tuiles rectangulaires de largeur 1). Ces résultats sont en cours d'écriture.

Depuis 2010, je fais également partie d'une action de Recherche Coopérative sur les Calculs par Systèmes Continus Complexes et Concurrents ($CaCO_3$), rassemblant des chercheurs de l'EPI Carte du LORIA, de membres du LIX et de l'équipe Escape, qui cherche à étudier sous un point de vue modèle de calcul les dynamiques de populations, ainsi qu'à étudier les petites machines de calcul à comportement complexe. Dans ce cadre, j'ai étudié les comportements calculatoires (degrés Turing) réalisables par pavages. J'ai caractérisé avec Pascal Vanier ces comportements, et montré en particulier que les pavages ne peuvent pas simuler certains comportements des machines de Turing [I1, U2].

Perspectives

Les perspectives de recherche s'inscrivent dans deux cadres : Les travaux sur le modèle des pavages, et les travaux sur les modèles de calcul en général. En ce qui concerne les pavages, j'envisage de travailler sur deux des principaux défis de la dynamique symbolique : la décidabilité de la conjugaison et la caractérisation des ensembles sofiques. Ensuite j'espère généraliser certaines des méthodes algorithmiques pour traiter d'autres modèles de calcul.

La question des invariants de conjugaison de la thèse de Pascal Vanier mérite d'être prolongée. En particulier nous nous intéressons maintenant à la décidabilité de la conjugaison en elle-même, ou plus exactement de sa place dans la hiérarchie arithmétique. La conjugaison dans $\mathbb{Z}$ (sur la droite) reste un important problème ouvert, et j'essaie actuellement de l'étudier sous une approche théorie des graphes, où la conjugaison s'exprime par une relation d'équivalence définie par transformations locales.

La caractérisation des ensembles de pavages sofiques est encore à ce jour difficile : on connaît quelques critères simples pour qu'un ensemble de coloriages soit réalisable par un jeu de tuiles, mais aucune condition suffisante utilisable. En particulier, nous avons introduit dans le mémoire plusieurs classes d'ensemble de pavages dont il est pour l'instant impossible de déterminer s'ils sont sofiques ou non. Le groupe de travail Pytheas Fogg auquel je participe occasionnellement va sans doute en particulier s'intéresser à cette question prochainement.

Nous réfléchissons actuellement à la création d'une ANR Jeunes-Chercheurs qui cherche à étudier la théorie des systèmes dynamiques sous l'angle de la calculabilité et de la théorie algorithmique de l'aléatoire. Plus particulièrement j'aimerais généraliser les résultats sur les comportements calculatoires des pavages à d'autres systèmes dynamiques plus concrets. Il est en effet fort probable que les méthodes utilisées restent pertinentes, et ainsi produire des systèmes dynamiques très simples (affines par morceaux) sans points récursifs.

Enfin, je compte proposer un sujet de thèse sur les fragments de la logique du premier ordre. On connaît par exemple plusieurs fragments logiques C (par exemple $C = \forall\exists\forall$) pour lesquels on sait démontrer que toute formule ϕ peut être transformée récursivement en une formule ψ de C en préservant la satisfiabilité. On se demande ici s'il est également possible de préserver non seulement la satisfiabilité, mais également le degré des théories étendant ϕ. Ainsi, il existe des formules ϕ tel que toute théorie complète étendant ϕ est non récursive. Peut-on trouver une telle formule ϕ de la classe $\forall\exists\forall$? Cette question est directement reliée aux objets que j'étudie, puisqu'en effet les différentes preuves de tels résultats sont liés à des codages dans ses fragments logiques de machines de Turing et de pavages. Pour obtenir ces généralisations, il serait nécessaire d'inventer des nouveaux codages, plus robustes.

INTRODUCTION

On trouvera ici un résumé de mes activités sur les pavages, c'est à dire principalement sur l'étude des coloriages du plan discret, comme vus par Wang [Wan61]. Les chapitres n'ont pas été écrits dans l'intention d'être encyclopédiques ou techniques, mais afin d'offrir une vision originale des pavages : on y trouvera ainsi à côtés des grands classiques des résultats inédits ou tout simplement oubliés.

Le domaine de pavages subit une résurgence dans ces dernières années grâce à l'intérêt des dynamiciens, pour lesquels la terminologie employée est plus souvent la notion de shift de type fini (SFT). L'indécidabilité de la pavabilité du plan a sursis leurs ambitions initiales [Lin04], mais on remarque désormais que la présence de la calculabilité pour l'étude des pavages n'est pas une contrainte mais un atout : Toutes les questions de nature dynamique ou combinatoire admettent souvent des réponses très précises, si tant est qu'on accepte que ces réponses soient de nature calculatoire [Sim11b, HM10, Mey10, Hoc09b, AS09]. Certains des résultats présentés ici vont même plus loin, et tendent à montrer que le modèle des pavage est bien plus qu'*un* modèle de calcul, et qu'il serait sans doute plus juste de le considérer comme une déclinaison de la machine de Turing plutôt qu'un modèle de part entière, tant il partage de propriétés avec celle-ci.

La deuxième approche qui sous-tend les résultats est la quête d'une caractérisation claire des espace de pavages sofiques. La question se pose ainsi : Quels sont les ensembles de coloriages du plan qu'on peut réaliser par (la projection d') un jeu de tuiles ? Les espaces de pavages sofiques jouent le rôle dans notre cadre des langages rationnels. Cependant, on ne dispose par pour l'instant de critères simples permettant de caractériser ces ensembles. Il n'existe pas d'équivalents des résiduels, pas d'équivalent de la congruence syntaxique, et tout au plus un analogue faible du lemme de l'étoile [Mat98a] et certaines conditions nécessaires [Pav11] ou suffisantes [AS, DRS10]. Certains des résultats présentés ici visent à mieux comprendre ces ensembles.

Ces deux approches vont être évoquées dans ce mémoire au travers de trois aspects bien différents. Le chapitre 2 s'intéresse aux bases, c'est à dire aux coloriages très particuliers qui doivent apparaître dans tout espace de pavages, donnant ainsi des conditions suffisantes de réalisabilité. On verra bien entendu que la plupart de ces conditions sont de nature calculatoire. Le chapitre 3 introduit des ensembles de coloriages très particuliers, et montre en particulier à quel point les pavages peuvent être compliqués, se voulant ainsi un contrepoint au chapitre précédent. Les constructions de ce chapitre codent en particulier des machines de Turing et expliquent ainsi pourquoi ce modèle de calcul est si proche du modèle originel. Enfin dans le chapitre 4 la problématique est étudiée du point de vue logique, point de vue qui historiquement a introduit les pavages comme présentés ici. On y cherchera en particulier à caractériser, comme en complexité descriptive, les ensembles de pavages réalisables par des fragments logiques.

Contributions

L'approche présentée ici est une des premières démarches profondément *structurantes*. Ce point de vue se manifeste de deux façons différentes. L'ensemble des pavages d'un jeu de tuiles ne sera pas vu uniquement comme un ensemble, mais on cherchera à le présenter comme une structure algébrique très particulière. Nous avons ainsi développé l'usage de l'ordre partiel induit par les motifs introduit par Durand, comme on l'examinera au chapitre 2.1.2. Voir les pavages comme modèles d'une théorie logique comme présenté au chapitre 4 a ainsi permis d'introduire de nouvelles classes d'espaces de pavages.

Une des originalités de la démarche présentée ici est également un refus systématique de considérer l'indécidabilité comme une fin en soi. Il est en effet relativement convenu que la plupart des problèmes de décision sur les pavages vont s'avérer algorithmiquement indécidables, mais ces résultats sont à la fois *imprécis* et *limités*.

La levée de cette imprécision est probablement le point le plus important guidant les recherches présentées ici. Ainsi, l'indécidabilité de la pavabilité du plan de façon périodique est en soi un résultat partiel : Il indique que les points périodiques d'un jeu de tuiles peuvent exhiber un comportement calculatoire, sans spécifier ce comportement. On révélera un lien avec la classe de complexité NP dans l'annexe E. De même, l'indécidabilité de la pavabilité du plan suggère également un résultat semblable concernant tous les pavages d'un jeu de tuiles, mais on verra au contraire au chapitre 2.2 que les espaces de pavages ne sont pas capables d'exhiber tous les comportements calculatoires d'une machine de Turing.

Le joug de l'indécidabilité conduit à s'intéresser uniquement à des espaces de pavages très particuliers, souvent avec des structures auto-similaires fortes et sans points périodiques. Ces outils sont donc limités et ne permettent pas d'étudier de façon fine tous les espaces de pavages. En particulier, les travaux présentés ici sont certainement parmi les premiers étudiant précisément les espaces de pavages contenant des points périodiques, et en particulier les espaces de pavages dénombrables. Dans ce contexte, les outils structurants développés ci-avant s'avèrent essentiels, et les résultats présentés au chapitre 3.3 et aux annexes C et F offrent une vision très claire. En particulier cette approche nécessite de nouvelles constructions, et le codage clairsemé du chapitre 3.3, entièrement original, en est un élément important.

Perspectives

Les outils développés ici ont permis une étude quasi complète des espaces de pavages contenant des points périodiques, au moins dans le cas dénombrable, et la réflexion structurante se porte maintenant sur les points (strictement) quasipériodiques. Pour en être capable il est nécessaire de mieux comprendre ces configurations particulières. On dispose à l'heure actuelle de nombreux modèles de jeux de tuiles exhibant des comportements quasipériodiques différents. De façon assez surprenante, on arrive dans la majorité de ces exemples à encoder le calcul d'une machine de Turing, mais aucune méthode *systématique* n'est connue. Une question essentielle apparaît ainsi : Peut-on encoder (une forme) du calcul dans tout jeu de tuiles apériodique ?

Une des motivations principales de la recherche de structure dans l'ensemble des pavages reste la question d'une caractérisation, ou tout au moins de conditions nécessaires fortes, pour qu'un ensemble de configurations soit réalisable par un jeu de tuiles, c'est à dire déterminer s'il s'agit d'un espace de pavages sofique. Les interrogations laissées ouvertes aux chapitres 1.2.2, 2.1.2 et 4.2 confirment la difficulté de cette question. En particulier la question posée dans la partie 4.2 sur l'existence des Π_2 non sofiques, pourtant d'énoncé simple, est une nouvelle preuve du seuil qu'il reste à combler. Nous espérons prolonger les approches développées ici afin de répondre à cete question.

Enfin une question naturelle est la généralisation des méthodes calculatoires aux autres modèles de calcul proches des pavages, comme les machines de Turing ou les automates cellulaires. Là encore, il s'agit de préciser les résultats d'indécidabilité obtenus. On sait ainsi qu'on ne peut décider s'il existe une configuration sur laquelle une machine de Turing ne s'arrête pas [Hoo66], mais on ne sait que depuis récemment [KO08] produire et étudier les machines de Turing sans configuration immortelle périodique. En particulier, on ne connait pas à l'heure actuelle de machine de Turing immortelle sans configuration immortelle récursive. Couplé au codage de Kari présenté au chapitre 3.4.2, une résolution de ce problème permettrait de donner une nouvelle preuve du résultat de Myers [Mye74], voire du résultat de Simpson [Sim11b]. Une situation similaire se profile pour les ensembles limites d'automates cellulaires (qui sont en particulier des espaces de pavages), encore mal compris.

Contenu

Le document est constitué de quatre chapitres, dont un chapitre d'introduction, et d'une annexe. L'annexe reproduit certains des articles écrits ces dernières années sur le sujet. On y a omis certaines contributions sur les pavages qui ne s'intégraient pas dans le cadre de ce document [I2] ou estimées redondantes [I5, A1].

Le premier chapitre introduit les diverses définitions utilisées dans la suite. La théorie des pavages faisant converger plusieurs domaines, la terminologie varie souvent d'un article à l'autre. Le vocabulaire utilisé ici est la notion d'espace de pavages, inspiré du livre de Sadun [Sad08], qui correspond exactement aux sous-décalages (sous-shifts) de la dynamique symbolique [LM95, Lin04]. Pour plus de clarté, on choisit ici de différencier symboliquement les deux notions : le terme sous-shift est réservé à la dimension 1, c'est à dire aux coloriages de la droite $\mathbb{Z}$, ie aux coloriages des mots biinfinis, tandis que le vocabulaire espace de pavages s'applique en dimension 2, pour les coloriages de $\mathbb{Z}$.

Les notions importantes sont introduites combinatoirement, i.e. par motifs interdits. Cette approche a été choisie au détriment de l'approche topologique (fermé invariant par décalage(s)) qui, bien qu'utilisée dans certaines démonstrations, est moins à même d'exprimer certains résultats. Les trois objects fondamentaux introduits sont les espaces de pavages de type fini (équivalents aux pavages par tuiles de Wang), les espaces sofiques et les espaces effectifs.

Le chapitre 2 regarde les espaces de pavages "de l'intérieur". On s'y intéresse aux *bases*, c'est à dire aux coloriages très particuliers qui doivent apparaître dans tout espace de pavages non vide. Ainsi par exemple tout tel ensemble contient un coloriage *quasipériodique*. On y décrira des théorèmes de bases géométriques (existence de points avec une direction de périodicité), combinatoires (avec peu de motifs différents) ou algorithmiques (calculatoirement peu compliqués). Certaines de ces résultats proviennent de théorèmes existants [Kre53, Sho60, JS72c] sur les ensembles effectifs (ensembles Π_1^0) dont les espaces de pavages sont un exemple important. Les autres théorèmes sont dans leur ensemble récents et fruits de l'école marseillaise [Dur99, I7, DLS08, I1].

Le chapitre 3 étudie les différents codages d'une machine de Turing par un jeu de tuiles. On y étudie bien entendu le codage bien connu et conventionnel utilisé dans la plupart des preuves de **NP**-complétude ou d'indécidabilité [Lew78, vEB97, Har85, LP98], ainsi que d'autres codages plus originaux, en particulier le codage clairsemé que nous avons introduits et qui permet de caractériser précisément les ensembles de degré Turing réalisables par pavages, ainsi que le codage encore peu compris introduit par Kari [Kar07] et le codage oublié de Aanderaa et Lewis [AL74, Lew79].

L'étude des théories logiques, et plus particulièrement du fragment $\forall\exists\forall$ de la logique du premier ordre, est fortement liée à l'etude des pavages, et ce n'est pas un hasard si la majorité des premiers théorèmes sur les pavags sont dus à des logiciens reconnus [Wan61, Han74, Rob71, KMW62, Büc62a, Her71, GK72, Mak74, Mye74, AL74, Poi80, Har85]. Nous étudierons cette connexion au chapitre 4, qui se veut un compagnon à l'article [I8]. On y étudie donc les codages classiques permettant de voir toute formule du fragment $\forall\exists\forall$ comme représentant un ensemble de pavages, et réciproquement comment tout espace de pavages peut se représenter comme une formule de ce fragment. On obtiendra ensuite quelques liens entre les espaces de pavages usuels (de type fini, sofiques) et certaines classes de formules de la logique MSO, généralisant ainsi la connexion sur les mots finis entre langages rationnels et MSO obtenue par Büchi et Elgot [Büc62b, Elg61].

PREMIER CHAPITRE

DÉFINITIONS

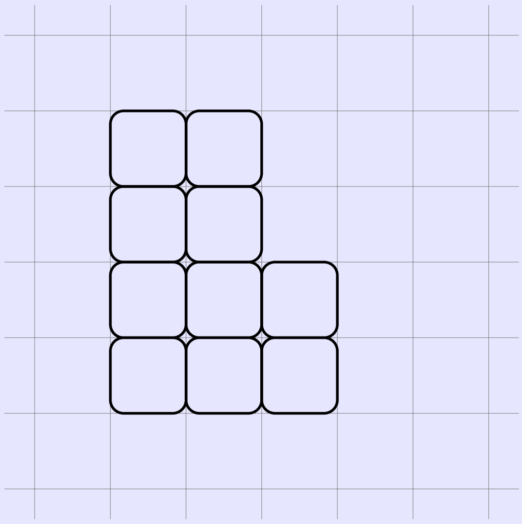
Nous commençons bien entendu par les définitions des objets étudiés. Bien que les pavages soient étudiés depuis maintenant plusieurs décennies, le vocabulaire n'est pas totalement figé. Le terme motif peut ainsi désigner aussi bien une partie finie du plan qu'un coloriage de cette partie. De même un pavage peut désigner un coloriage du plan ou un coloriage *valide* du plan. Nous nous tiendrons dans la suite aux définitions données ici. Le terme *espace de pavages* est inspiré de [Sad08].

1.1 Motifs

Commençons par définir ce qu'est un motif, c'est à dire un coloriage d'une partie finie du plan. Les définitions sont relativement naturelles, et il n'y a que peu à en dire.

Définition 1.1 — *Formes*

Une *forme* est une partie de $\mathbb{Z}^2$.

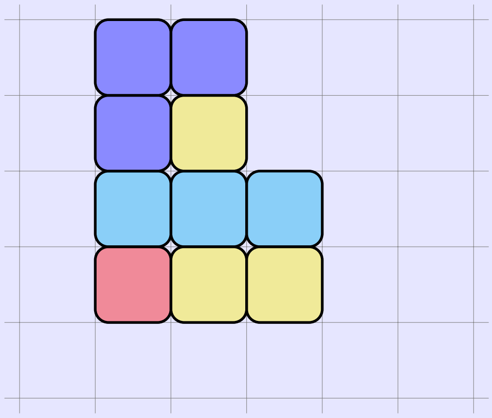


Si I est une forme et t un vecteur de $\mathbb{Z}^2$, on note $I + t = \{x + t, x \in I\}$

On s'intéresse principalement ici à deux types de formes : les formes finies (I est un ensemble fini) et la forme $I = \mathbb{Z}^2$, le plan tout entier.

Définition 1.2 *Motif*

Soit C un ensemble (souvent fini) dit de couleurs. Un *motif* M de C est une application d'une forme I dans C. La forme I est appelé support du motif.



Si $J \subseteq I$, on note $\mathsf{M}_{|J}$ la restriction de M à J, qui est donc un motif de support J.

Si M est un motif, et t un vecteur de $\mathbb{Z}^2$, on note $\mathsf{M} + t$ le motif de support $I + t$ défini par $\forall x, (\mathsf{M} + t)(x) = \mathsf{M}(x - t)$.

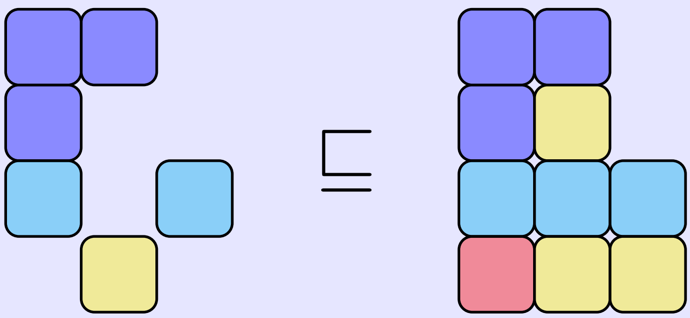
On dit que deux motifs M et N sont équivalents, ce qu'on note $M \sim N$ s'il existe t tel que $\mathsf{M} = \mathsf{N} + t$.

Notons que dans les dessins, il n'est pas précisé où se situe l'origine du plan $\mathbb{Z}^2$. On travaillera en effet souvent à translation près : On assimilera le motif M et le motif $\mathsf{M} + t$.

Définition 1.3

Si M et N sont deux motifs de support I et J, on dira que M est inclus dans N ou encore que M est un sous-motif de N, ce qu'on note $\mathsf{M} \sqsubseteq \mathsf{N}$ s'il existe t tel que $I + t \subseteq J$ et $\mathsf{M} + t = \mathsf{N}_{|I+t}$

Dit autrement, il existe K tel que $\mathsf{N}_{|K} \sim \mathsf{M}$. On dit alors dans ce cas que M apparaît dans N en position K, ou que M est facteur de N, ou encore que N contient M en position K.



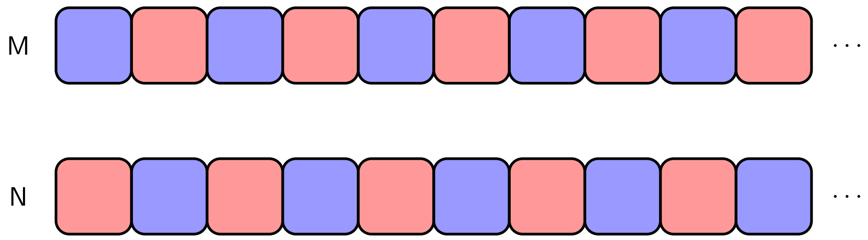


FIGURE 1.1 – Deux motifs infinis (correspondant aux mots $(01)^\omega$ et $(10)^\omega$) qui vérifient $\mathsf{M} \sqsubseteq \mathsf{N}$ et $\mathsf{N} \sqsubseteq \mathsf{M}$ mais qui ne sont pas équivalents.

Pour les motifs finis (ce qui nous intéresse principalement), si $\mathsf{M} \sqsubseteq \mathsf{N}$ et $\mathsf{N} \sqsubseteq \mathsf{M}$, alors $\mathsf{N} \sim \mathsf{M}$. Ce n'est pas vrai en général. Il suffit pour s'en convaincre de considérer les deux demi-droites (ie motifs définis par exemple sur $\mathbb{N} \times \{0\}$) en figure 1.1

1.2 Pavages

Dans toute la suite, et sauf indication contraire, on ne s'intéressera qu'aux motifs *finis*. L'idée essentielle qu'il s'agit de capturer ici est la *localité*. On ne s'intéresse qu'aux propriétés des coloriages du plan qu'on peut tester localement, c'est à dire en observant le coloriage sur des zones finies (mais potentiellement arbitrairement grandes). La définition suivante est donc naturelle :

Définition 1.4 *Configuration*

Une *configuration* est un motif de support $\mathbb{Z}^2$. On note $C^{\mathbb{Z}^2}$ l'ensemble des configurations sur l'ensemble de couleurs C, et $C^{[\mathbb{Z}^2]}$ l'ensemble des motifs finis sur l'ensemble C. On omettra l'indice C lorsque l'ensemble de couleurs, comme souvent dans la suite, sera fixé.

Si X est une configuration, on note $\mathcal{L}(\mathsf{X})$ l'ensemble des motifs finis que contient X.

$$\mathcal{L}(\mathsf{X}) = \left\{ \mathsf{M} \in C^{[\mathbb{Z}^2]} \middle| \mathsf{M} \sqsubseteq \mathsf{X} \right\}$$

Si I est une forme, on note $\mathcal{L}_I(\mathsf{X})$ l'ensemble des motifs finis de support I que contient X, et $p_I(\mathsf{X})$ le cardinal de $\mathcal{L}_I(\mathsf{X})$.

On cherche donc à manipuler les configurations à travers l'ensemble des motifs qu'elles contiennent ; autrement dit la seule information utile sur X est $\mathcal{L}(\mathsf{X})$.

La notion d'espace de pavages est donc définie en suivant ce principe :

Définition 1.5 *Espaces de pavage*

Soit C un ensemble de couleurs.
Soit U un ensemble de configurations et V un ensemble de motifs finis. On définit

$$\begin{array}{lcllcl}
\mathcal{L}(U) & \triangleq & \{\mathsf{M} \in C^{[\mathbb{Z}^2]} | & \exists \mathsf{X} \in U, \mathsf{M} \sqsubseteq \mathsf{X}\} & = & \bigcup_{\mathsf{X}\in U} \mathcal{L}(\mathsf{X}) \\
\mathcal{F}(U) & \triangleq & \{\mathsf{M} \in C^{[\mathbb{Z}^2]} | & \forall \mathsf{X} \in U, \mathsf{M} \not\sqsubseteq \mathsf{X}\} & = & {}^c\mathcal{L}(U) \\
\mathcal{S}(V) & \triangleq & \{\,\mathsf{X} \in C^{\mathbb{Z}^2} \,| & \forall \mathsf{M} \in V, \mathsf{M} \not\sqsubseteq \mathsf{X}\} & = & \{\mathsf{X} \in C^{\mathbb{Z}^2} | \mathcal{L}(\mathsf{X}) \cap V = \emptyset\}
\end{array}$$

$\mathcal{F}(U)$ est l'ensemble des motifs finis qui n'apparaissent dans aucune des configurations de U. $\mathcal{S}(V)$ est l'ensemble des configurations qui ne contiennent aucun des motifs de V. $\mathcal{L}(U)$ est également appelé *langage* de U. Un ensemble S de configurations est appelé un *espace de pavages* (ou encore TS pour Tiling Space) s'il est égal à $\mathcal{S}(V)$ pour un ensemble de motifs V. Si $S \subseteq S'$ sont deux espaces de pavages, on dira que S est un sous-espace de pavages de S'.

S est dit de type fini (TSFT pour Tiling Space of Finite Type) s'il est égal à $\mathcal{S}(V)$ pour un V *fini*.

Un espace de pavages en dimension 1 (les configurations sont de support $\mathbb{Z}$) est aussi appelé un *sous-shift*.

$\mathcal{F}$ et $\mathcal{S}$ forment la connexion de Galois correspondant à la relation $\not\sqsubseteq$, ce qui signifie principalement :

$$V \subseteq \mathcal{F}(U) \iff \mathcal{S}(V) \supseteq U$$

V est souvent appelé *motifs interdits*, puisqu'il s'agit des motifs qui ne peuvent apparaître dans aucune des configurations de $\mathcal{S}(V)$. Le vocabulaire sous-shift utilisé en dimension 1 tient ses racines dans la théorie de la dynamique topologique.

Si S est un espace de pavages, on appellera *pavage* toute configuration de S. L'idée est de distinguer un pavage (configuration vérifiant les contraintes) d'une configuration (configuration qui ne vérifie a priori rien de particulier).

D'une façon un peu plus concrète que la définition, un espace de pavages est donné par un certain nombre de *contraintes locales*, chacune représentée par un motif $\mathsf{M} \in V$. Une contrainte locale peut par exemple représenter le fait que tout point de couleur verte possède à sa droite un point de couleur rouge. On l'exprimera en ajoutant à notre ensemble V tous les motifs 2×1 dans lequel le point de gauche est vert, mais le point de droite n'est pas rouge.

L'attrait fondamental des pavages est vraiment cette notion de localité : Le fait qu'une configuration est un pavage ne dépend pas de la configuration elle-même, mais plus précisément des motifs qu'elle contient.

En particulier :

Fait 1.1

Si X est un pavage, et $\mathcal{L}(\mathsf{Y}) \subseteq \mathcal{L}(\mathsf{X})$ alors Y est un pavage.

Plus généralement, si X_i sont des pavages et $\mathcal{L}(\mathsf{Y}) \subseteq \cup_i \mathcal{L}(\mathsf{X}_i)$ alors Y est un pavage.

Cette dernière relation est en fait une caractérisation :

Fait 1.2

S est un espace de pavages si et seulement si

$$\forall (X_i)_{i\in I} \in S, \forall Y, \left[\mathcal{L}(\mathsf{Y}) \subseteq \bigcup_{i\in I} \mathcal{L}(\mathsf{X}_i)\right] \implies Y \in S$$

En particulier, considérons un ensemble de contraintes V de sorte que l'espace de pavages correspondant à V contienne un pavage X avec un (seul) point de couleur rouge et tous les autres de couleur vert. Alors la configuration Y toute verte vérifie $\mathcal{L}(\mathsf{Y}) \subseteq \mathcal{L}(\mathsf{X})$. Donc c'est un pavage. Ainsi, *on ne peut pas forcer l'apparition d'une couleur* avec des pavages.

Les caractérisations ci-haut permettent très facilement de montrer qu'un ensemble de configuration est ou non un espace de pavages. On verra qu'il est souvent plus délicat de prouver qu'un espace de pavages n'est pas un TSFT, ou un sofique.

1.2.1 Type fini

Un TSFT correspond au cas où il y a un nombre *fini* de contraintes locales. Dans ce cas, on peut quelquefois trouver plus naturel de représenter l'espace de pavages par des motifs *autorisés* : On se donne alors une forme (finie) I, et une liste U de motifs de support I. Une configuration X est alors dans l'espace de pavages défini par U si et seulement si tout motif de support I présent dans X est un motif de U. Dit autrement, pour savoir si une configuration X est dans l'espace de pavage, il suffit de promener une fenêtre de forme I sur la configuration. Si elle ne rencontre que des motifs autorisés, la configuration est bien un pavage, sinon ce n'est pas le cas.

Comme premier exemple, considérons les contraintes représentées sur la figure 1.2. On cherche donc les configurations X telles que tous les carrés 2×2 de X fassent partie de l'ensemble des motifs indiqués sur la figure. Un raisonnement rapide montre qu'il n'y a que trois configurations possibles :

- Un échiquier infini rouge et blanc (ce qui fait deux configurations, suivant si le point en position $(0,0)$ est rouge ou blanc) ;
- La configuration toute verte.

Notons en particulier qu'aucune des configurations obtenues ne contient le troisième motif.

On peut obtenir une caractérisation des TSFT similaire à celle obtenue précédemment pour les espaces de pavages généraux :

Fait 1.3

S est un TSFT si et seulement si il existe une forme F tel que

$$\forall (X_i)_{i\in I} \in S, \forall Y, \mathcal{L}_F(\mathsf{Y}) \subseteq \bigcup_{i\in I} \mathcal{L}_F(\mathsf{X}_i) \implies Y \in S$$

On rappelle que $\mathcal{L}_F(\mathsf{X})$ est l'ensemble des motifs M de support F présents dans X.

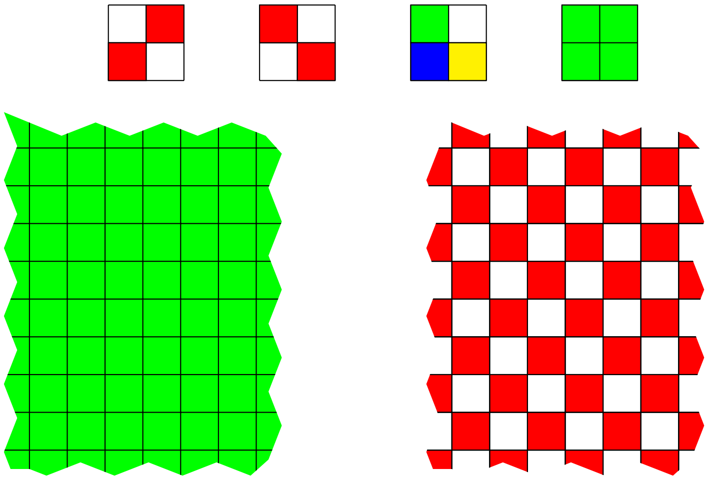


FIGURE 1.2 – Contraintes locales (motifs autorisés) et les pavages obtenus

En particulier, reprenons l'exemple d'un TSFT S qui contient un pavage X avec un (seul) point de couleur rouge et tous les autres de couleur vert. Prenons la forme F correspondant à ce TSFT. Considérons une configuration Y avec deux points de couleur rouge suffisamment éloignés (c'est à dire qu'aucun motif de forme F ne peut les contenir simultanément). Alors cette configuration vérifie $\mathcal{L}_F(\mathsf{Y}) \subseteq \mathcal{L}_F(\mathsf{X})$. Donc c'est un pavage. Ainsi, *on ne peut pas forcer une couleur à apparaître au plus une fois* avec un nombre fini de contraintes.

Tuiles de Wang

Une façon commune de définir des pavages sur $\mathbb{Z}^2$ est la notion de *tuiles de Wang* [Wan61]. Soit Q un ensemble de couleurs. Une tuile de Wang est un quadruplet de Q, qu'on représente comme sur la figure 1.3. Etant donné un ensemble de tuiles de Wang τ, un *pavage* par τ associe à chaque élément de $\mathbb{Z}^2$ une tuile de τ de la façon suivante : une tuile A ne peut être placée à la droite d'une autre tuile B que si la couleur de droite de la tuile B est égal à la couleur de gauche de la tuile A. Une règle similaire est observée pour placer une tuile au dessus d'une autre (voir figure 1.4). On notera $\mathcal{S}(\tau)$ l'espace de pavages ainsi défini.

L'espace de pavages ainsi obtenu est bien un TSFT : Il suffit de prendre comme ensemble de couleurs l'ensemble des tuiles de τ. Les motifs interdits de V sont alors les motifs de taille 2×1 et 1×2 correspondant à des tuiles de Wang mal collées. Il est de même assez aisé de transformer un TSFT en un jeu de tuiles de Wang. Ici, transformer signifie que les pavages engendrés par chacun des modèles sont essentiellement les mêmes (la notion exacte est la *conjugaison*, que nous étudierons rapidement en fin de chapitre).

La figure 1.5 contient un premier exemple. A noter que les tuiles de Wang correspondent bien au concept ludo-éducatif tant rencontré des pièces de puzzle . On trouvera ainsi en figure 1.6 des pièces de puzzle correspondant aux tuiles de la figure 1.5

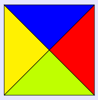


FIGURE 1.3 – Une tuile de Wang

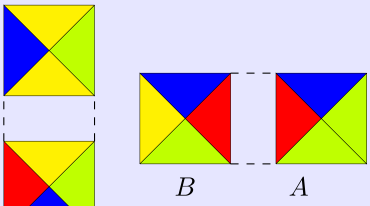
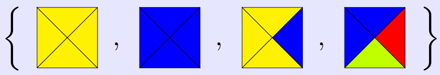
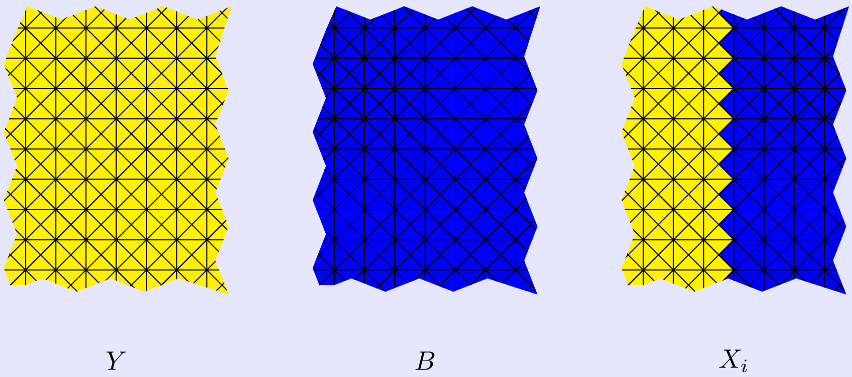


FIGURE 1.4 – Règles de placement des tuiles de Wang

Tuiles

$\tau =$ 

Pavages obtenus



En utilisant uniquement les tuiles de τ, on obtient 3 types de pavages :

- Un pavage tout jaune Y
- Un pavage tout bleu B
- Des pavages X_i à moitié bleu et à moitié jaune. les X_i ne diffèrent les un des autres que par la position de la frontière entre la zone bleue et la zone jaune, indexée par un entier $i \in \mathbb{Z}$; il n'y a donc qu'un seul pavage à décalage près.

FIGURE 1.5 – Un exemple de pavages par tuiles de Wang.

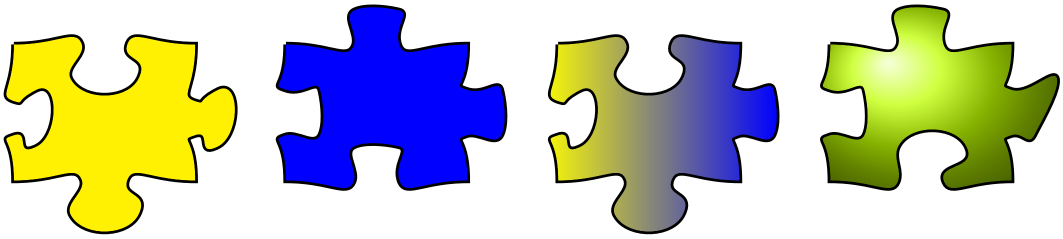


FIGURE 1.6 – Représentation sous forme de pièces de puzzle des tuiles de Wang.

1.2.2 Sofiques

Définitions

Soient C_1 et C_2 deux ensembles (finis) de couleurs et π une application de C_1 dans C_2. π se relève en une application (qu'on nommera toujours π) de l'ensemble $C_1^{\mathbb{Z}^2}$ des configurations sur C_1 dans l'ensemble $C_2^{\mathbb{Z}^2}$ des configurations sur C_2 en appliquant π composante par composante, et de même en une application des motifs $C_1^{[\mathbb{Z}^2]}$ sur C_1 dans les motifs $C_2^{[\mathbb{Z}^2]}$ sur C_2. On notera logiquement $\pi(\mathsf{M})$ l'image de M par π.

Définition 1.6

Un ensemble S est un espace de pavages sofique si et seulement si il existe un TSFT S' et π tel que $S = \pi(S')$.

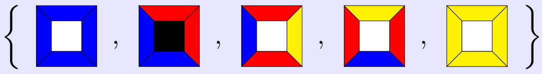
Autrement dit, un ensemble sofique est un TSFT dans lequel on a confondu certaines couleurs.

Vue la définition, il n'est pas clair qu'un ensemble sofique soit vraiment un espace de pavages. En général, si S est un espace de pavages, alors $\pi(S)$ aussi : Les motifs interdits de $\pi(S)$ sont les motifs M pour lesquels $\pi^{-1}(\mathsf{M}) \supseteq \mathcal{F}(S)$ (Aucun antécédent du motif M n'apparaît dans une configuration de S). Ce résultat est cependant valide uniquement si le nombre de couleurs $|C_1|$ est fini, la preuve faisant intervenir le théorème de compacité (voir fin de ce chapitre).

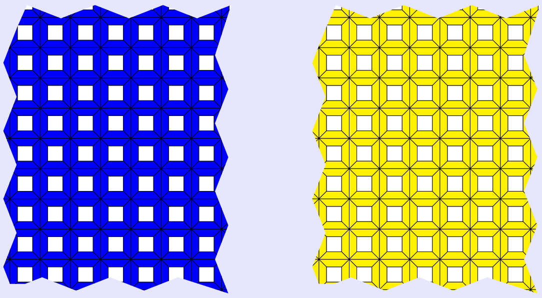
On peut définir également les espaces de pavage sofiques à partir du formalisme des tuiles de Wang. Il suffit de considérer des tuiles de Wang contenant une couleur supplémentaire, comme représentées sur la figure 1.7. On définira alors une tuile de Wang décorée comme un élément de $Q^4 \times C_2$. L'espace de pavages sofique correspondant à un jeu de tuiles décorées τ est alors obtenu en prenant les pavages du plan par les tuiles de τ et en ne retenant que la couleur au centre. On obtient bien ainsi un sous-ensemble de ${C_2}^{\mathbb{Z}^2}$.

L'exemple de la figure 1.7 est un exemple très important qui reviendra sous une forme ou une autre dans la suite. Le jeu de tuiles τ produit un TSFT qui permet de "marquer" une cellule particulière de l'espace $\mathbb{Z}^2$, la couleur jaune représentant les cellules au nord est de cette cellule distinguée, la couleur bleue les autres. Si on oublie les couleurs en ne gardant que ce qui est au centre, on obtient un espace de pavages contenant uniquement deux configurations (à translation près) : la configuration toute blanche, et la configuration avec exactement un point noir. Ceci est l'exemple classique d'un espace de pavages sofique qui n'est pas un TSFT, comme indiqué précédemment dans le fait 1.3.

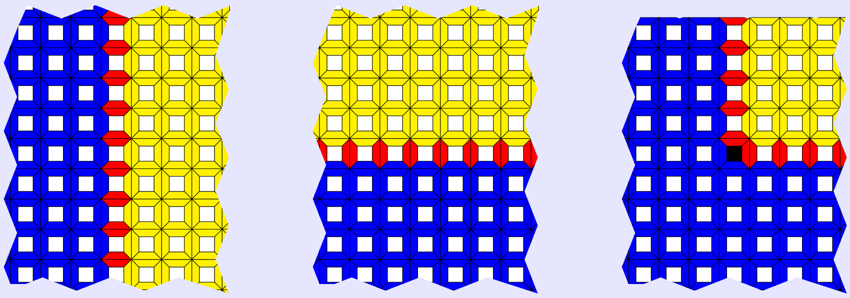
Tuiles

$\tau =$ 

Pavages (avant projection)

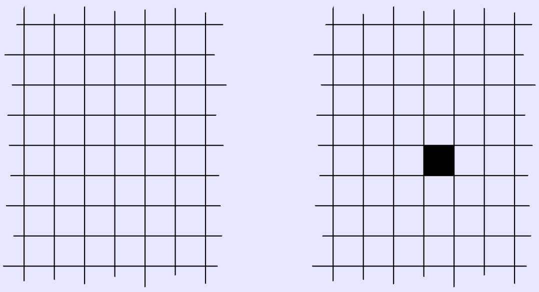


B Y



V_i H_j $C_{i,j}$

Pavages (après projection)



W N_{ij}

FIGURE 1.7 – Exemple d'un espace de pavages sofiques. τ est un jeu de tuiles de Wang décorées, qui génère l'espace de pavages $\{B, Y, V_i, H_j, C_{ij}, (i,j) \in \mathbb{Z}\}$. Lorsqu'on projette pour ne garder que la couleur au centre, on obtient l'espace de pavages sofique $\{W, N_{ij}, (i,j) \in \mathbb{Z}\}$.

Conditions nécessaires

Il est difficile de trouver une caractérisation combinatoire des espaces de pavages sofiques, semblable à celle obtenue dans les faits 1.2 et 1.3 pour les espaces de pavages respectivement quelconque et de type fini. Si on s'intéresse aux espaces sofiques en dimension 1 (coloriage de $\mathbb{Z}$, autrement dit un langage de mots biinfinis), tout se simplifie : un espace de pavage est sofique si et seulement si son langage (l'ensemble des mots finis qu'il contient) est rationnel. La notion de langage rationnel est plus délicate dans notre cas (voir en particulier [GR97]), ce qui empêche une caractérisation précise.

On peut obtenir un résultat partiel, en étendant le concept de *fooling set* des langages rationnels dans notre cadre :

Définition 1.7

Soit F une forme et X, Y deux configurations. On note $\mathsf{X} \ltimes_F \mathsf{Y}$ la configuration égale à X sur F et à Y sur son complémentaire.

Soit U un espace de pavages et F une forme. Un ensemble $S \subseteq U$ de configurations est un ensemble *trompeur* pour U de support F si pour tout $\mathsf{X}, \mathsf{Y} \in S$, soit $\mathsf{X} \ltimes_F \mathsf{Y} \notin U$ soit $\mathsf{Y} \ltimes_F \mathsf{X} \notin U$.

Proposition 1.4

Si U est un espace sofique, alors il existe une constante δ tel que tout ensemble trompeur pour U de support $[\![0, n]\!] \times [\![0, m]\!]$ soit de taille bornée par $\delta^{2(n+m)}$.

L'idée de la preuve est simple. Supposons U sofique, obtenu comme projection par π d'un TSFT U' qu'on peut supposer être donné par tuiles de Wang. Notons δ le nombre de tuiles de Wang. Prenons maintenant un ensemble $S = \{(\mathsf{X}_i)_{i \in I}\}$ trompeur de support $F = [\![0, m]\!] \times [\![0, n]\!]$.

A chaque élément X_i de S correspond (au moins) un élément Y_i de U' tel que $\pi(\mathsf{Y}_i) = \mathsf{X}_i$. Regardons la forme $R = [\![0, n]\!] \times [\![0, m]\!] \setminus [\![1, n-1]\!] \times [\![1, m-1]\!]$, qui représente le *contour* du rectangle F.

Si deux pavages Y et Y' coincident sur R, $\mathsf{Y} \ltimes_F \mathsf{Y}'$ est aussi un pavage, puisque les contraintes des tuiles de Wang ne concernent que des cellules adjacentes.

Ainsi si Y_i et Y_j coincident sur R, on en déduit alors que $\mathsf{Y}_i \ltimes_F \mathsf{Y}_j$ est un pavage de U', donc que $\mathsf{X}_i \ltimes_F \mathsf{X}_j$ est un pavage de U, de même que $\mathsf{X}_j \ltimes_F \mathsf{X}_i$, ce qui implique $i = j$. On en déduit donc que tous les Y_i diffèrent sur R, et donc qu'il y a au plus $\delta^{|R|}$ éléments dans S.

A noter que cette condition n'est pas suffisante. En particulier prenons un espace de pavages U pour lequel le nombre de motifs de taille $n \times m$ est (au plus) simplement exponentiel en $n + m$. Il vérifiera donc trivialement les conditions de la proposition. Cependant rien pour autant n'indique qu'il soit sofique. Trouver un espace de pavages vérifiant cette condition (ainsi qu'une autre condition nécessaire de nature calculatoire qu'on examinera au chapitre suivant) et qui ne soit pas sofique reste un intéressant problème ouvert.

1.2.3 Effectifs

Dans les deux cas précédents, il est facile, par un argument de compacité, d'énumérer pour un espace de pavages (de type fini ou sofique) S l'ensemble des motifs interdits $\mathcal{F}(S)$: Un motif M est interdit dans un TSFT (resp un sofique) S s'il existe une forme I telle qu'aucun motif valide de forme I ne contienne M (resp. M n'apparaît dans la projection d'aucun motif valide de forme I). Ainsi dans les deux cas, l'ensemble $\mathcal{F}(S)$ est donc énumérable algorithmiquement. Ces deux exemples se situent donc dans une classe plus large d'espaces de pavages, les effectifs :

Définition 1.8

Un espace de pavages S est *effectif* s'il existe un algorithme qui énumère $\mathcal{F}(S)$. De façon équivalente, il existe un ensemble V récursif de motifs interdits tel que $S = \mathcal{S}(V)$.

Pour construire V récursif, il suffit de partir d'une énumération de $\mathcal{F}(S)$ et de la rendre croissante en remarquant qu'on peut toujours remplacer un motif M de support I par tous les motifs de support $J \supseteq I$ qui contiennent M.

On trouvera à la figure 1.8 un exemple d'espace de pavages effectif qui, par une application simple de la proposition 1.4 n'est pas sofique.

On peut bien sûr généraliser cette notion, et parler, pour un oracle A, d'espace de pavages A-effectif s'il existe une énumération récursive en A de motifs interdits.

Lorsqu'on étudie des propriétés de calculabilité, les espaces de pavages naturels à étudier en dimension 1 sont les espaces effectifs [Mil11, CDK08], et ils jouent le rôle des TSFT en dimension 2. En effet, en général, toute propriété de nature calculatoire vraie sur un TSFT en dimension 2 sera vraie sur un espace effectif en dimension 1, et réciproquement. Il ne s'agit pas ici d'un théorème, mais uniquement d'un principe, qui s'explique par le théorème suivant :

Théorème 1.5 (*[AS, DRS10]*)

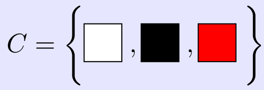
Soit S un espace de pavages en 1D sur l'ensemble de couleurs C.

On note $S^{(2)}$ l'espace de pavages en 2D constitué de configurations où toutes les lignes sont identiques, égales à un élément X de S.

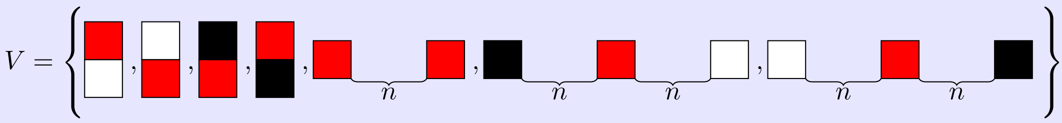
Alors S est effectif si et seulement si $S^{(2)}$ est sofique.

Ce théorème permet en particulier de construire facilement des sofiques, voire des TSFT, avec des propriétés très particulières : Il suffit de le construire en 1D avec des effectifs, et d'utiliser ce théorème pour "relever" le résultat en 2D.

Couleurs

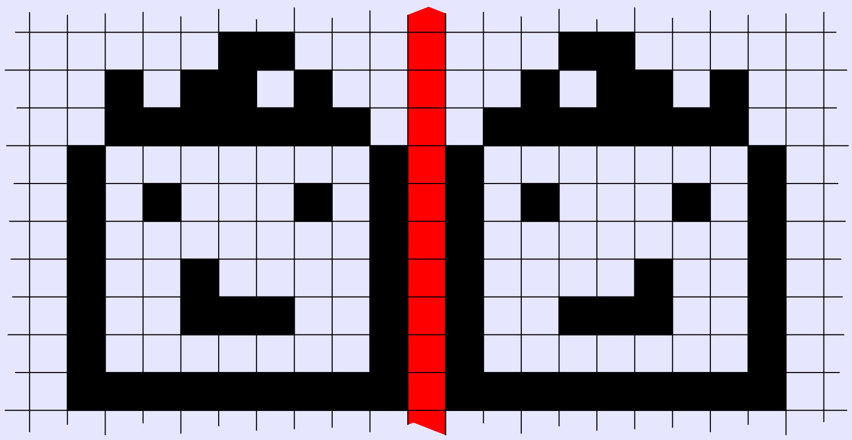


Motifs interdits



Les quatre premiers motifs obligent chaque colonne à être uniquement constituée de rouge ou de noirs et blancs. Les motifs en cinquième position imposent au plus une seule colonne rouge. Les motifs en sixième et septième position obligent la configuration à être symétrique par rapport à la colonne rouge (si celle-ci existe).

Pavages obtenus



On peut bien entendu obtenir également des pavages sans colonne rouge (autrement dit n'importe quelle configuration sur les couleurs noires et blanches)

FIGURE 1.8 – Exemple d'un espace de pavages effectif qui n'est pas sofique

1.3 Topologie

Bien que nous ayons donné ici des définitions combinatoires des espaces de pavage, il sera cependant souvent plus utile de les voir comme des objets de la dynamique symbolique, dans lesquels ils sont généralement appelés sous-shifts. Nous réservons cependant dans ce manuscrit le terme sous-shift aux espaces de pavages unidimensionels.

L'idée est de considérer l'ensemble des configurations $C^{\mathbb{Z}^2}$ comme un espace topologique muni de la topologie produit.

On définira alors les espaces de pavages ainsi :

Définition 1.9

S est un espace de pavages si et seulement si il est fermé et invariant par translation ($\mathsf{X} \in S, t \in \mathbb{Z}^2 \implies \mathsf{X} + t \in S$).

Il est assez facile de vérifier que ces deux définitions sont équivalentes. L'avantage de la définition topologique est qu'elle fait apparaître de façon claire le fait que S est compact. En particulier une intersection d'espaces de pavages est vide si et seulement si une intersection finie est vide, les projetés de TSFT (ie les espaces sofiques) sont bien des espaces de pavages (comme images d'un compact par une application continue, etc).

On peut même y définir les TSFT ainsi (voir par ex. [Sch00]) : Un espace de pavages S est un TSFT si pour tout ensemble I et toute suite $(S_i)_{i\in I}$ d'espaces de pavages, $\cap_{i\in I} S_i = S$ implique qu'il existe J fini tel que $\cap_{j\in J} S_j = S$

Bien que le vocabulaire et les notions de dynamique symbolique ne seront que peu évoqués dans ce manuscrit, ils sont essentiels à bien des preuves et des motivations. On pourra en particulier lire [Sch00, LM95, Lin04] pour une approche dynamique.

Une des questions centrales en dynamique symbolique est de comprendre quand deux espaces de pavages X et Y sont “isomorphes” (plus exactement conjugués), c'est à dire quand il existe une bijection continue commutant avec les translations entre X et Y. Une manière d'approcher ce problème est via la notion d'*invariant de conjugaison* : Il s'agit de quantités définies pour tout espace de pavages (par exemple leur cardinalité) et qui sont égales pour deux espaces conjugués. La plupart des résultats qu'on obtient ici, en particulier dans les deux prochains chapitres, peuvent s'interpréter en termes d'invariants de conjugaison.

DEUXIÈME CHAPITRE

BASES

Une première manière de comprendre les espaces de pavages est d'essayer de se demander à quoi ressemblent les configurations d'un espace de pavages. En particulier, peut-on y trouver des configurations vérifiant des propriétés particulières ?

Définition 2.1 *Base*

On dit qu'une propriété P (donc un ensemble de configurations) est une *base* pour une famille d'espaces de pavages S si tout espace de pavages de S non vide possède une configuration vérifiant la propriété P.

On va s'intéresser dans la suite aux différents théorèmes de base qu'on peut obtenir sur les espaces de pavages. On en trouvera de plusieurs types, selon qu'ils s'appliquent uniquement aux espaces de pavages de type finis, ou s'ils s'appliquent à des espaces plus généraux. On s'intéresse également aux théorèmes d'anti-base, qui expriment le fait qu'on peut trouver des espaces de pavages où la propriété P n'est jamais vraie.

2.1 Combinatoire

2.1.1 Minimaux

Un des tout premiers théorèmes obtenus dans la théorie des pavages est un théorème d'anti-base : il existe un TSFT non vide sans point périodique (une configuration X de support $\mathbb{Z}^2$ est périodique si elle admet d vecteurs de périodicité non colinéaires, ou, de façon équivalente, si l'ensemble de ses translatés $\{\mathsf{X}+t, t \in \mathbb{Z}^2\}$ est fini). Ce théorème obtenu par Berger [Ber64] est une première étape vers la démonstration qu'il est impossible algorithmiquement de décider si un TSFT (donné, par exemple, par un jeu de tuiles de Wang) est vide ou non. Nous en reparlerons plus loin.

Notons qu'on sait construire depuis longtemps des espaces de pavages, même unidimensionels (i.e. des sous-shifts), non vides et sans point périodique : Le sous-shift de $\{0,1\}^{\mathbb{Z}}$ où on interdit tous les facteurs $auaua$ (où a est une lettre) est sans point périodique, et il est non vide, puisqu'il contient par exemple le mot de Thue-Morse [Thu12, Mor21].

Ce théorème d'anti-base a été plus tard transformé en un théorème de base. Pour cela introduisons la notation d'espace minimal :

Définition 2.2

Un espace de pavages est minimal s'il est non vide et ne contient aucun sous-espace de pavages strict (non vide).

Le sous-shift décrit ci-haut interdisant les motifs $auaua$ est un exemple non trivial de sous-shift minimal [GH64].

A quoi ressemblent les configurations d'un espace de pavages minimal ?

Définition 2.3

Une configuration X est dite *minimale/quasipériodique/uniformément récurrente* si pour tout motif M, il existe une fenêtre (un support) J tel que M apparaisse dans tout motif de X de support J.

Les configurations périodiques sont en particulier quasipériodiques. Le mot de Thue-Morse est un exemple de configuration quasipériodique non périodique [Mor21]. La notion de points quasipériodiques a été introduite par Birkhoff [Bir12], qui en dit :

> Parmi les mouvements d'un système dynamique, il peut en exister certains ayant la propriété remarquable de représenter le mouvement tout entier avec tel degré d'approximation qu'on désire, pendant *tout* intervalle de temps égal à T, où T varie seulement avec le degré de l'approximation. [...] ces mouvements sont appelés mouvements [uniformément] récurrents ; ils forment une extension naturelle des mouvements périodiques.

Il est assez facile de voir qu'une configuration est minimale si et seulement si elle appartient à un espace de pavages minimal.

On peut alors démontrer que tout espace de pavages contient un espace de pavages minimal [Bir12], ce qui signifie :

Théorème 2.1 (***Les points quasipériodiques forment une base [Dur99]***)

Tout espace de pavages non vide contient une configuration quasipériodique.

Ce théorème sous cette forme est attribué à Durand [Dur99], mais peut se lire déjà dans l'article de Birkhoff.

2.1.2 Ordre

La preuve de Durand utilise une relation de préordre. En examinant de près cette relation, on peut obtenir d'autres théorèmes de base. C'est ce que nous avons fait dans [I7], reproduit dans l'annexe C. La relation est la suivante :

Définition 2.4

On dit que $\mathsf{X} \prec \mathsf{Y}$ si $\mathcal{L}(X) \subsetneq \mathcal{L}(Y)$.
On note $\mathsf{X} \preceq \mathsf{Y}$ si $\mathcal{L}(X) \subseteq \mathcal{L}(Y)$.

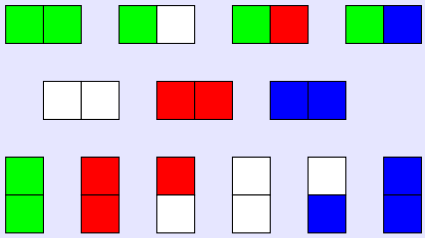
Cette relation peut également être appelée relation de récurrence en dynamique topologique : $\mathsf{X} \preceq \mathsf{Y}$ si X est dans l'adhérence de l'orbite de Y. Autrement dit, tout espace de pavages contenant Y contient X.

On trouvera un exemple du diagramme de Hasse d'un TSFT bien choisi sur la figure 2.1.

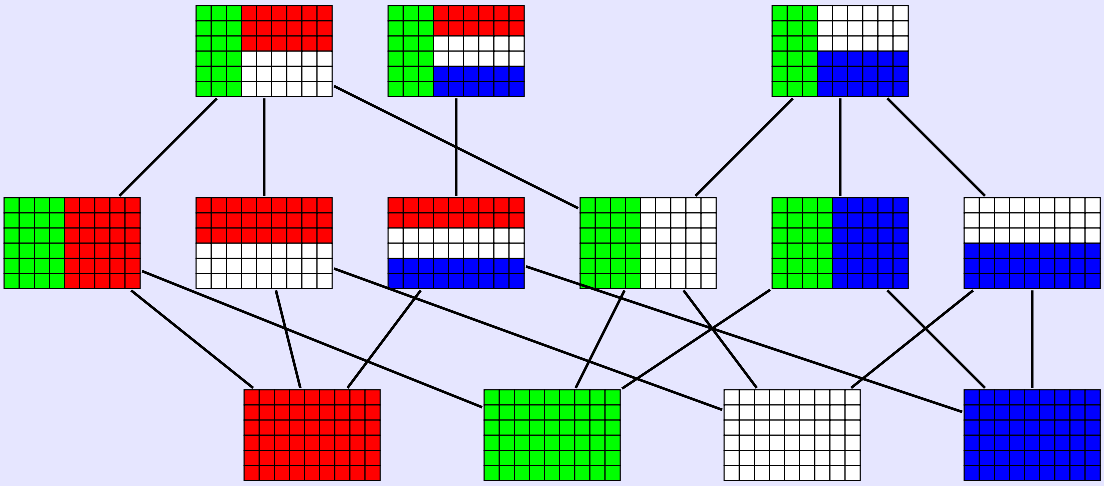
L'existence de configurations quasipériodiques s'obtient dans [Dur99] en montrant qu'il existe des configurations minimales pour l'ordre ou, mieux, en montrant que sous chaque configuration Y il existe une configuration minimale $\mathsf{X} \preceq \mathsf{Y}$.

On peut utiliser ce préordre, qui structure l'ensemble des pavages d'un espace de pavages, pour obtenir d'autres résultats. Ainsi, tout espace de pavages S contient par exemple un pavage maximal (dans S) pour $\preceq$ (Théorème C.3). Cependant il est assez difficile de comprendre ce que

Motifs autorisés *(Tous les dominos non dessinés sont interdits)*



Pavages obtenus



Représentation symbolique

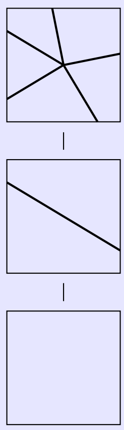


FIGURE 2.1 – Le diagramme de Hasse d'un espace de pavages bien choisi. On remarque en particulier que tous les pavages au niveau 0 sont (quasi)périodiques, et que les pavages au niveau 1 ont une (seule) direction de périodicité. On peut représenter symboliquement (et prouver qu'ils sont ainsi) les pavages de la façon suivante : les pavages au niveau 0 par un fond uniforme, les pavages au niveau 1 par une droite qui sépare le plan en deux (l'épaisseur de la droite peut être arbitrairement grande, voir par exemple le drapeau du Luxembourg), et ceux au niveau 2 par des demi-droites qui s'intersectent. Les couleurs ont été choisies de façon à représenter le drapeau de la république d'Afrique du Sud (1856-1877).

représentent ces pavages maximaux, d'autant plus qu'il s'agit d'éléments maximaux dans S et non pas dans l'ensemble des configurations.

Une perspective plus intéressante est de regarder le bas de l'ordre : Une configuration X est au niveau 1 si et seulement si toute configuration inférieure à X est minimale (donc au niveau 0). On peut alors démontrer que l'exemple de la figure n'est pas un hasard, mais en fait un théorème :

Théorème 2.2 (***Les 1-périodiques forment une base dans le cas dénombrable)***)

Si S est un TSFT infini *dénombrable*, alors S contient une configuration X ayant exactement une direction de périodicité.

Si S est un espace de pavages infini dénombrable, alors S contient
- Soit une configuration X ayant exactement une direction de périodicité
- Soit une configuration X périodique partout sauf sur un motif fini.

On en trouvera une preuve dans l'annexe C. L'idée est de montrer, dans le cas dénombrable, que le niveau 1 existe (c'est à dire que S ne peut pas contenir uniquement des configurations périodiques (Théorème C.8), et qu'il n'existe pas de chaîne infinie décroissante convergeant vers des configurations périodiques) puis de le caractériser.

A noter qu'un théorème similaire dans le cas non dénombrable est délicat à obtenir et ce pour plusieurs raisons : D'abord on peut avoir un espace de pavages S non dénombrable constitué uniquement de points minimaux (typiquement des configurations quasipériodiques non périodiques). De plus, même s'il ne contient pas de configurations quasipériodiques non périodiques, le niveau 1 peut être bien plus compliqué que la caractérisation ci-dessus, en tout cas dans le cas général. Ainsi, avec un peu de réflexion, on peut construire un espace de pavages S non dénombrable, avec un seul point minimal périodique, qui ne rentre pas dans la classification ci-dessus.

Nous conjecturons cependant que ce n'est pas possible avec un TSFT :

Conjecture 1

Si S est un TSFT infini, alors il contient soit une configuration quasipériodique non périodique, soit deux configurations périodiques distinctes.

Sans vouloir rentrer dans les détails, cette conjecture permettrait de prouver qu'il n'existe pas d'automate cellulaire non nilpotent, mais nilpotent sur toute configuration quasipériodique ou encore qu'un automate cellulaire est nilpotent si et seulement si son ensemble limite est uniquement ergodique.

2.1.3 Complexité

La complexité d'une configuration X est la fonction qui à un support J associe le nombre de motifs distincts présents dans X de support J. En général, on étudiera plus souvent la fonction $p_n(X)$ qui compte le nombre de motifs distincts de support un carré de taille n. La fonction p_n est une première mesure de la complexité d'une configuration en fonction de ses motifs. Une manière plus précise de regarder est à travers la complexité de Kolmogorov : On dira qu'une configuration X est de complexité (bornée par) f si pour tout motif M présent dans X de taille n, $K(\mathsf{M}|n) \leq f(n) + O(1)$.

Ces deux quantités sont bien entendu reliées : Si une configuration est de complexité de Kolmogorov f alors elle a moins de $2^{f(n)+O(1)}$ motifs de taille n. Réciproquement si l'ensemble $\mathcal{L}(\mathsf{X})$ des motifs de X est calculable, alors la complexité de Kolmogorov de X est bornée par $\log p_n(X)$ (la complexité de Kolmogorov peut cependant être plus élevée si $\mathcal{L}(\mathsf{X})$ n'est pas calculable).

Un théorème de base a été obtenu sur la complexité de Kolmogorov :

Théorème 2.3 (*Les configurations de complexité sous-linéaire forment une base [DLS08]*)

Soit S un TSFT non vide. Alors il existe un pavage X et une constante c tels que pour tout n et tout motif M de X de support $[1;n]^2$, on a $K(\mathsf{M}) \leq c.n + O(1)$

n devient n^{d-1} en dimension $d > 1$. Il s'agit bien d'une complexité sous-linéaire puisque ces motifs sont constitués de n^2 (resp. n^d en dimension d) cellules.

L'idée de la preuve est la suivante : Pour chaque n, il est facile d'exhiber un motif de support $[1;n]^2$ qui soit de complexité $c.n$: Il suffit de décrire un contour valide (ce qui nécessite $4 \log |C| n$ bits) et un algorithme pour le remplir, par exemple un algorithme exponentiel. Cependant les motifs à l'intérieur de celui-ci peuvent potentiellement être complexes. On utilise alors une ingénieuse construction récursive pour forcer tous les motifs de taille inférieure à $n/2$ qu'il contient à être également de faible complexité. On peut alors conclure par compacité à l'existence de notre pavage.

On peut généraliser ce théorème de plusieurs façons. D'abord en remarquant qu'il s'applique également aux espaces sofiques (la projection d'un pavage "simple" est toujours "simple"), et ensuite en remarquant qu'il s'applique à tous les motifs et non pas uniquement aux motifs carrés. Si I est une forme, notons $\delta(I)$, la frontière de I, les points de I qui ont un voisin non dans I :

$$\delta(I) = \{z \in I | B(z,1) \not\subseteq I\}$$

On peut alors écrire :

Corollaire 2.4

Soit S un espace de pavages sofique. Alors il existe un pavage X et une constante c tels que pour toute forme I et tout motif M de X de support I, on a $K(\mathsf{M}|I) \leq c.|\delta(I)| + O(1)$.

En particulier, il existe un pavage pour lequel le nombre de motifs distincts de support I est borné par $2^{O(|\delta(I)|)}$.

Il suffit pour prouver le corollaire de décomposer I en composantes 8-connexes et de remarquer que chacune de ces composantes J est incluse dans un carré de côté au plus $|\delta(J)|$ et enfin d'utiliser le théorème sur chacun de ces carrés.

Remarquons que ce théorème ne peut pas se généraliser aux espaces de pavages en général : Le fait qu'on puisse construire un motif valide uniquement à partir de son contour est une vraie spécificité. En particulier :

Théorème 2.5

Pour tout $\rho < 1$, il existe un espace de pavages effectif S tel que dans tout pavage X de S, tout motif M de taille n vérifie $K(\mathsf{M}) \geq \rho n^2 - O(1)$. De plus chaque pavage vérifie $p_n(X) = 2^{\Omega(n^2)}$.

(Notons en particulier que tout sous-espace de pavages minimal inclus dans S est d'entropie positive.)

Preuve : On considère l'espace de pavages effectif où on interdit tout motif M de taille n vérifiant $K(\mathsf{M}) \leq \rho n^2$ pour n suffisament grand. On peut démontrer, par exemple par un argument type lemme local de Lovász, que l'espace de pavages obtenu est non vide [RU06].

Si maintenant X est une configuration, il est clair que pour tout k, et tout motif M de X de taille n, on a $K(\mathsf{M}) \leq k^2 \log |C| p_k(X) + \frac{n^2}{k^2} \lceil \log_2 p_k(X) \rceil$. A k constant, cela implique, pour n suffisament grand, $\lceil \log_2 p_k(X) \rceil \geq \rho k^2$. ■

2.2 Calculabilité

On va maintenant s'intéresser à ce qu'on peut dire de la complexité algorithmique des pavages d'un espace de pavages. Un pavage, élément de $C^{\mathbb{Z}^2}$, peut être vu comme une fonction de $\mathbb{Z}^2$ dans C, ou encore comme le ruban (l'oracle) d'une machine de Turing. Ces deux définitions donnent le même objet. On dira ainsi qu'un pavage est récursif si sa fonction est récursive, ou s'il existe une machine de Turing qui écrit sur son ruban de sortie chaque cellule du pavage dans l'ordre.

Le premier théorème d'antibase obtenu est celui de Myers [Mye74], qu'on examinera au chapitre suivant, et qui mentionne qu'il existe des TSFT non vides sans point récursif. Cependant, comme on va le voir ici, tout TSFT contient toujours des points relativement "simples".

Rappelons d'abord très brièvement la notion de degré Turing : Si X et Y sont deux pavages (resp. langages, mots infinis, etc), on dira que $X \leq_T Y$ s'il existe une machine de Turing qui avec oracle Y et sur l'entrée n calcule la couleur de la n-ème cellule de X (resp. décide si $n \in X$, calcule la n-ème lettre de X, etc) On notera $X \equiv_T Y$ si $X \leq_T Y$ et $Y \leq_T X$. Le degré Turing d'un pavage (langage, etc) X est sa classe d'équivalence pour la relation $\equiv_T$, qu'on notera $\deg_T X$. On notera enfin 0 le degré des points récursifs, et $0'$ le degré de l'arrêt.

On peut maintenant prouver :

Théorème 2.6 (*Les configurations $\boldsymbol{\Delta^0_2}$ forment une base*)

Tout espace de pavages effectif S contient une configuration minimale de degré Turing inférieur ou égal à $0'$.

Preuve : Soit $\mathcal{F}$ l'ensemble de motifs interdits définissant S, et soit $(\mathsf{M}_n)_{n\in\mathbb{N}}$ une énumération (calculable) de tous les motifs (valides ou non).

On définit récursivement $\mathcal{F}_n$ de la façon suivante : $\mathcal{F}_{-1} = \emptyset$ et $\mathcal{F}_{n+1} = \mathcal{F}_n \cup \{\mathsf{M}_{n+1}\}$ si l'espace de pavages défini par $\mathcal{F}_n \cup \{\mathsf{M}_{n+1}\} \cup \mathcal{F}$ est non vide, et $\mathcal{F}_{n+1} = \mathcal{F}_n$ sinon.

Il est clair que $\tilde{\mathcal{F}} = \cup_n \mathcal{F}_n$ est calculable avec oracle l'arrêt, puisqu'il suffit de savoir décider, étant donné un ensemble de motifs interdits (donné par exemple par une machine de Turing), s'il engendre un espace de pavages vide ou non.

Maintenant, il est clair, par construction, que $\tilde{\mathcal{F}}$ définit un espace de pavages $\tilde{S}$ minimal. De plus son complémentaire L est exactement l'ensemble des motifs valides de $\tilde{S}$.

Maintenant, il est clair qu'il existe un pavage X calculable en L : Il suffit de le définir récursivement en $n \in \mathbb{N}$: on pose $\mathsf{M}_0 = \epsilon$ (le motif vide) et on définit M_{n+1} comme le plus petit motif (dans l'ordre lexicographique) de L de support $[-(n+1), n+1]$ ayant M_n en son centre. Il est alors clair que $\mathsf{X} = \cup_n \mathsf{M}_n$ convient. ■

Notons qu'il s'agit d'un théorème en deux parties : tout espace de pavages effectif contient un espace de pavages minimal dont le langage est de degré inférieur à $0'$, et tout espace de pavages minimal contient un pavage récursif en son langage. On reviendra sur cette remarque plus tard.

En fait ce théorème est une variante d'un théorème de base dû à Kreisel sur les ensembles effectifs, notion qu'on va maintenant définir :

Définition 2.5 *Ensemble Π^0_1*

Soit $S \subseteq \{0,1\}^{\mathbb{N}}$. On dit que S est un ensemble Π^0_1 (plus exactement est une classe Π^0_1 d'ensembles) s'il existe une machine de Turing M telle que S soit exactement l'ensemble des oracles sur lesquels la machine de Turing ne s'arrête pas.

On voit beaucoup mieux l'intérêt dans notre cadre de cette notion avec la définition équivalente suivante :

Définition 2.6

Soit $S \subseteq \{0,1\}^{\mathbb{N}}$. S est un ensemble Π^0_1 s'il existe un langage $L \subseteq \{0,1\}^\star$ récursivement énumérable tel que S est l'ensemble des mots infinis ne contenant aucun *préfixe* dans L

La différence essentielle entre notre objet d'études et les ensembles Π^0_1 est donc que ces derniers interdisent des *préfixes*, alors que nous interdisons dans les espaces de pavages des *facteurs* (motifs).

On en déduit donc que les TSFT et, plus généralement, les espaces de pavages effectifs sont des cas particuliers d'ensembles Π^0_1.

En particulier, on obtient ainsi plusieurs théorèmes de base :

Théorème 2.7 (***Les hyperimmune-free, les bas, ... sont des bases***)

Soit S un espace de pavages effectif non vide. Alors

- S contient un pavage X dont le degré Turing est inférieur à $0'$ [Kre53]
- S contient un pavage X dont le degré Turing est strictement inférieur à $0'$ [Sho60]
- S contient un pavage X qui est hyperimmune-free [JS72c]
- S contient un pavage X dont le degré Turing est bas (low) [JS72c]
- S contient deux éléments X, Y dont les degrés Turing sont de minimum zéro [JS72c]

Nous ne détaillons pas tout de suite le vocabulaire, nous le ferons au besoin. On notera qu'une démonstration de l'existence d'un degré bas (low) dans le contexte des pavages peut se trouver dans [DLS08].

Peut-on améliorer (voire utiliser) ces théorèmes pour les espaces de pavages ? Par exemple, peut-on montrer qu'il existe un pavage à la fois minimal et de degré bas ? un pavage à la fois minimal et hyperimmune-free ?

Le théorème 2.6 présenté plus haut est précisément l'analogue dans notre cadre du théorème de Kreisel, et dans ce cas on peut même choisir ce pavage minimal. On ne sait pas en revanche si les autres théorèmes sont toujours valables en prenant des pavages minimaux.

Mais les espaces de pavages ne sont pas les seuls Π^0_1 qu'on rencontre : Soit S un espace de pavages effectif, donné par un ensemble de motifs interdits $\mathcal{F}$. Alors un langage L est exactement l'ensemble des motifs valides d'un sous-espace de S si et seulement si :

- L ne contient aucun motif de $\mathcal{F}$
- L contient un motif de taille 1
- Tout sous-motif d'un motif de L est dans L
- Tout motif de L est prolongeable en un motif de L plus grand le contenant en son centre

Toutes ces conditions sont effectives : Il existe bien une machine de Turing qui, étant donné un langage L en oracle, s'arrête si et seulement si L n'est pas le langage d'un sous-espace de S. Autrement dit, l'ensemble des sous-espaces de S est également un ensemble Π^0_1. En particulier :

Théorème 2.8 (*Once more, with feeling*)

Soit S un espace de pavages effectif non vide. Alors

- S contient un sous-espace $\tilde{S}$ dont le langage est de degré Turing inférieur à $0'$ [Kre53]
- S contient un sous-espace $\tilde{S}$ dont le langage est de degré Turing est strictement inférieur à $0'$ [Sho60]
- S contient un sous-espace $\tilde{S}$ dont le langage est hyperimmune-free [JS72c]
- S contient un sous-espace $\tilde{S}$ dont le langage est de degré Turing bas (low) [JS72c]

Et maintenant, on sait dire mieux : Le premier résultat se généralise, on peut donc supposer que $\tilde{S}$ est minimal (voir la preuve du théorème 2.6), mais les autres ne se généralisent pas :

On pourra trouver dans l'annexe D un sous-shift effectif (donc en 1D) et un TSFT (en 2D) dont tout sous-espace minimal S a un langage de degré au moins $0'$. Cet exemple ne répond cependant pas aux interrogations précédentes, puisque celui-ci contient des configurations minimales récursives. On touche ici du doigt une remarque essentielle : la complexité d'une configuration et la complexité de son language peuvent être très différentes.

Pour finir, notons que tous ces théorèmes sont uniquement la traduction dans le monde des espaces de pavages des théorèmes sur les ensembles Π^0_1. Nous avons démontré un théorème spécifique (voir annexe F), faux pour les ensembles Π^0_1 généraux :

Théorème 2.9

Tout espace de pavages (effectif ou non) non vide contient soit un point récursif (en fait périodique), soit deux points de degrés Turing différents mais comparables.

A l'inverse, Jockusch et Soare [JS72c] ont construit un ensemble Π^0_1 dont tous les points sont de degré Turing différents et incomparables.

Le théorème se prouve ainsi : tout espace de pavages non vide contient une configuration minimale X. Si X est périodique, elle est récursive. Sinon, elle est strictement quasipériodique. On peut alors montrer qu'on peut, récursivement en X, construire une infinité d'autres points minimaux. Plus exactement, on peut construire une machine de Turing qui étant donné X et un élément $y \in \{0,1\}^{\mathbb{N}}$ construit (injectivement) un autre point minimal $f(\mathsf{X}, y)$ de sorte qu'on puisse reconstituer y connaissant $f(\mathsf{X}, y)$.

Plus exactement

Théorème 2.10

Si X est un point minimal non récursif d'un espace de pavages S, alors S contient des pavages (minimaux) de degré Turing d pour tout $d \geq_T \deg_T \mathsf{X}$

Couplé au théorème 2.6, on en déduit ainsi :

Théorème 2.11

Tout espace de pavages effectif non vide contient soit un pavage périodique (donc récursif), soit des pavages de degré Turing d pour tout $d \geq_T 0'$.

On conjecture un théorème plus fort

Conjecture 2

L'ensemble des degrés Turing de tout espace de pavages sans point périodique est clos par le haut.

Il suffirait, pour prouver la conjecture, de montrer que si S est un espace de pavages, et $\mathsf{X} \in S$ un pavage de degré Turing d, alors il existe dans S un pavage minimal de degré inférieur à X. Nous n'avons cependant aucune idée sur une quelconque preuve de cette conjecture.

TROISIÈME CHAPITRE

CODAGE

Nous avons vu dans le chapitre précédent que tout TSFT contient des pavages avec des propriétés très particulières. Afin de savoir à quel point ces théorèmes sont précis, il est nécessaire de construire des ensembles de pavages les plus complexes possibles. Qu'il s'agisse de propriétés combinatoires ou calculatoires, l'outil principal est le codage de machines de Turing.

On recense de nombreux codages dans la littérature ; cependant on peut les répertorier en plusieurs familles, correspondant tout autant aux propriétés du codage qu'aux théorèmes qu'ils permettent de démontrer. Nous en donnons ici quatre exemples.

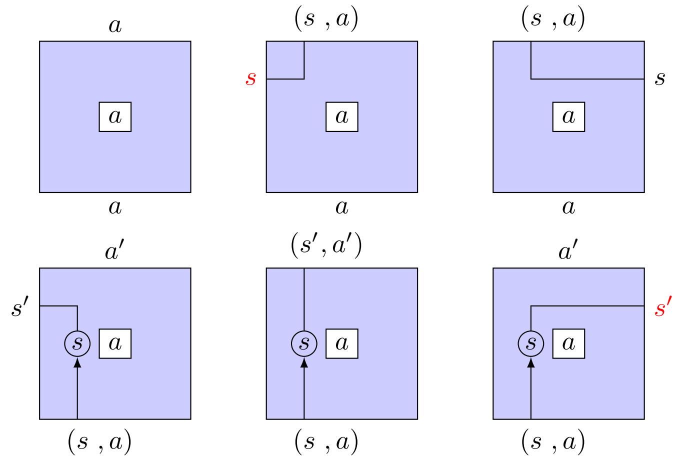
Codage d'une machine de Turing

Le codage d'une machine de Turing par un TSFT est simple et relativement standard. On en trouve de nombreuses variations dans la littérature, souvent présentées par le formalisme des tuiles de Wang. Des exemples visuels en sont en particulier présentés dès les premiers articles [KMW62, Wan63, Her71, Rob71].

Le codage s'appuie principalement sur les 6 (familles de) tuiles de Wang présentées figure 3.1. On rappelle qu'une configuration $c = (w, s, i)$ d'une machine de Turing est donnée par un mot biinfini $w \in \Sigma^{\mathbb{Z}}$, un état $s \in S$, et une position $i \in \mathbb{Z}$. Notons alors $T(c)$ le mot où on remplace la i-ème lettre de w par (s, w_i), obtenant ainsi un mot biinfini sur l'alphabet $(S \times \Sigma) \cup \Sigma$. Le codage a alors la propriété suivante : *Sur toute ligne bien pavée contenant $T(c)$ sur sa face sud, la face nord contient $T(c')$ où c' est l'évolution de la machine de Turing en une étape partant de c.* Réciproquement, sur toute ligne bien pavée contenant $T(c')$ sur sa face nord, la face sud contient $T(c)$ où c est une configuration qui évolue vers c'. En particulier, dans un demi-plan correctement pavé contenant une configuration c donnée en face sud, on observe donc le diagramme espace-temps de la machine de Turing partant de cette configuration.

A noter que rien ne contraint une face nord à contenir un mot de la forme $T(c)$: En particulier elle peut contenir un mot sur l'alphabet Σ (sans tête Turing) ou avec plusieurs symboles dans $S \times \Sigma$) (avec plusieurs têtes). Il est très facile d'empêcher avec des contraintes de pavages l'apparition de plusieurs têtes sur la même ligne. Cela sera cependant rarement nécessaire, comme on le verra dans la suite. En revanche, il est *impossible* de forcer l'apparition d'une tête par des contraintes locales.

Un problème similaire se pose avec la question de la configuration initiale : Dans un pavage valide d'un demi-plan, rien ne permet de s'assurer que la face sud contient bien une certaine configuration, et donc que l'évolution observée sur le demi-plan correspond bien à l'évolution de la machine de Turing à partir de sa configuration initiale. Ce problème est de même nature que le



Il y a une copie de la première tuile par lettre $a \in \Sigma$ de l'alphabet du ruban. Les deux tuiles suivantes existent pour toute paire $(s, a) \in S \times \Sigma$ d'état et de lettre. Enfin, les trois tuiles suivantes existent lorsque la lecture de a dans l'état s écrit a' et passe dans l'état s' en déplaçant la tête respectivement à gauche, sur place, et à droite.

FIGURE 3.1 – Codage d'une machine de Turing par tuiles de Wang.

problème précédent : il faut réussir à *forcer l'apparition* d'une tuile particulier, plus exactement d'une tuile contenant l'état initial dans ce cas précis.

Nous allons donc voir dans la suite plusieurs façons de résoudre ce problème : soit en s'intéressant à des variantes de la pavabilité du plan tout entier, soit en transformant les diagrammes espace-temps.

3.1 Codage exact

Le codage ci-avant est central dans la théorie des pavages, et peut-être utilisé tel quel pour prouver de nombreux théorèmes.

Le premier théorème obtenu directement est bien entendu :

Théorème 3.1 (*[Büc62a, KMW62]*)

Le problème suivant est indécidable :

NOM :	PAV-1
ENTRÉE :	Un jeu de tuiles de Wang τ et une tuile $t \in \tau$
SORTIE :	Est-ce qu'il existe un pavage du plan par τ contenant la tuile t ?

La preuve est très simple. Partons d'une machine de Turing M et constituons un jeu de tuiles en prenant les tuiles de la figure 3.1 et ajoutons les tuiles de la figure 3.2.

Examinons les pavages du plan contenant la première de ces nouvelles tuiles. Il est alors évident que la ligne infinie contenant cette tuile code le mot $T(c_0)$ où c_0 est la configuration initiale de la machine de Turing, et que les lignes en dessous de celle-ci sont remplies par la tuile blanche. Il existe alors un pavage du plan contenant cette tuile si et seulement si l'évolution de la machine de Turing à partir de c_0 *ne termine pas*, ce qui termine la preuve.

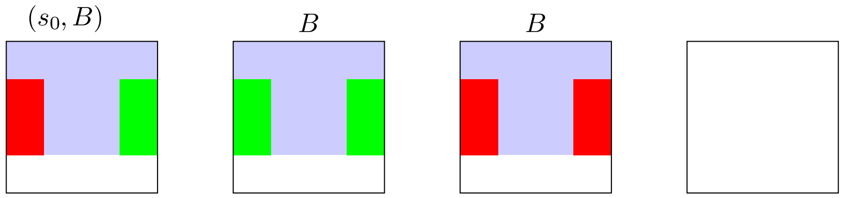


FIGURE 3.2 – Tuiles supplémentaire pour l'indécidabilité. s_0 désigne l'état initial et B le symbole blanc de la machine de Turing

Cette méthode de preuve permet en fait de montrer un résultat bien plus fort. Changeons la deuxième tuile de la figure 3.2 de sorte qu'elle puisse porter en face nord deux valeurs différentes, 0 ou 1. Pour le nouveau jeu de tuiles, chaque pavage correspond alors à une exécution infinie d'une machine de Turing partant d'une configuration initiale sur l'alphabet $\{0, 1\}$.

Si on se rappelle des ensembles Π_1^0 vu précédemment, on a donc démontré :

Théorème 3.2 (*[Han74, Sim11b]*)

> Pour tout ensemble Π_1^0 S, il existe un jeu de tuiles τ et une tuile $t \in \tau$ tels que l'ensemble des pavages par τ avec t au centre soit en bijection avec S.

Plus exactement, la bijection est un homeomorphisme récursif, voir [I1]. Ce théorème est implicitement contenu dans Hanf [Han74], qui en déduit le corollaire suivant. Prenons l'ensemble S des mots infinis w sur l'alphabet $\{0, 1\}$ tels que $w_i = 0$ (resp. $w_i = 1$) si la i-ème machine de Turing s'arrête sur l'entrée vide en ayant écrit 0 (resp. 1) sur son ruban. Dans les autres cas, il n'y a aucune contrainte sur la valeur de w_i. Il est clair que S est Π_1^0 : considérer la machine de Turing avec oracle w qui exécute en parallèle toutes les machines de Turing et qui vérifie, si jamais la $i$ème s'arrête, que la réponse est conforme à la valeur de w_i. Il est également clair que S ne contient *aucun point récursif* (plus exactement, l'ensemble des machines s'arrêtant en répondant 0 et l'ensemble des machines s'arrêtant en répondant 1 sont récursivement inséparables). On en déduit ainsi :

Corollaire 3.3 (*[Han74]*)

> Il existe un jeu de tuiles τ et une tuile $t \in \tau$ tel que l'ensemble des pavages par τ contenant t est non-vide et ne contient aucun point récursif.

On peut utiliser le même codage pour s'intéresser aux pavages des zones *finies* du plan : on impose la présence de certaines tuiles de Wang très spécifiques sur le bord d'un carré, et on peut alors se demander s'il existe un remplissage valide du carré. On peut résoudre ce problème en ajoutant au premier jeu de la figure 3.1 les tuiles de la figure 3.3. Si on cherche alors un pavage d'un rectangle de taille $s \times t$ en forçant les tuiles de la figure 3.3 à apparaître sur le bord, on s'aperçoit vite qu'un tel pavage existe si et seulement si la machine de Turing atteint son état final à partir de la configuration initiale en exactement t pas et en ayant utilisé un espace inférieur ou égal à s.

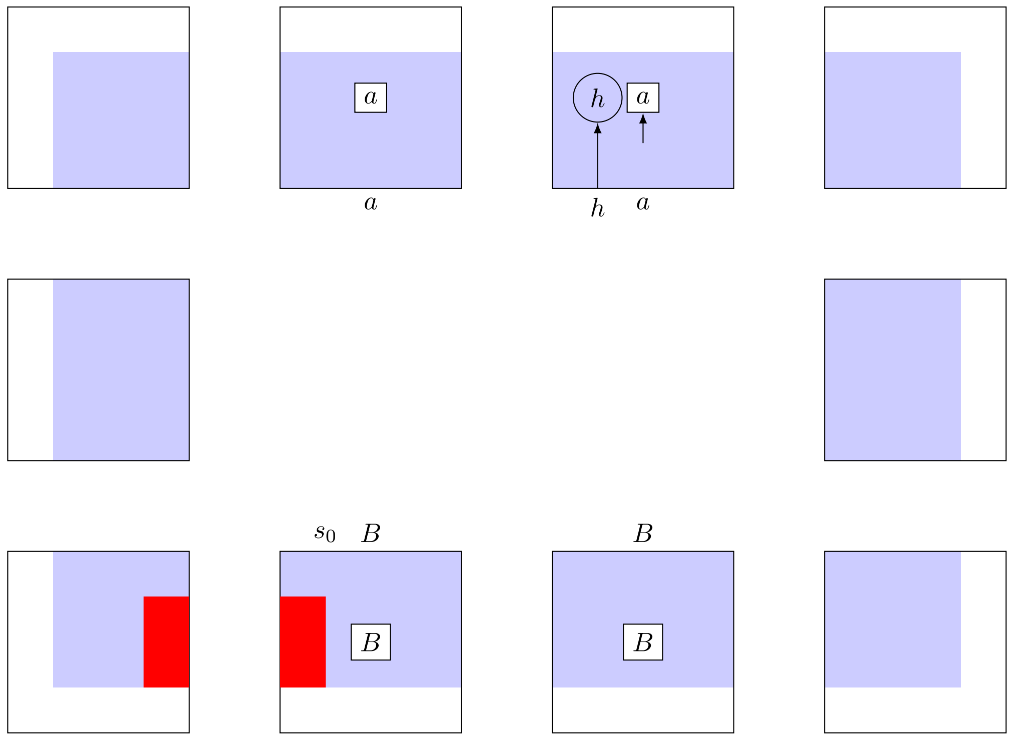


FIGURE 3.3 – Tuiles supplémentaires utilisées pour le pavage de zones finies du plan. s_0 désigne l'état initial, h l'état final et B le symbole blanc de la machine de Turing

On en déduit :

Théorème 3.4

Le problème suivant est indécidable :

NOM :	PAV-FIN
ENTRÉE :	Un jeu de tuiles τ et des tuiles $T \subset \tau$
SORTIE :	Est-ce qu'il existe un pavage d'un rectangle du plan par τ contenant les tuiles de T sur les bords ?

Si on code une machine de Turing non-déterministe, on obtient un résultat de *complexité* :

Théorème 3.5

Le problème suivant est NP-complet :

NOM :	PAV-FIN2
ENTRÉE :	Un jeu de tuiles τ, des tuiles $T \subset \tau$, et des entiers s,t écrits en unaire
SORTIE :	Est-ce qu'il existe un pavage d'un rectangle de taille $s \times t$ par τ contenant les tuiles de T sur les bords ?

En progressant un peu, on obtient une version plus forte :

Théorème 3.6

Le problème suivant est NP-complet :

NOM :	PAV-FIN3
ENTRÉE :	Un jeu de tuiles τ et des entiers s,t écrits en unaire
SORTIE :	Est-ce qu'il existe un pavage d'un rectangle de taille $s \times t$ par τ ?

La réduction à partir de PAV-FIN2 est simple : On construit d'abord un jeu constitué de st tuiles indexées par un élément de $[\![1, s]\!] \times [\![1, t]\!]$ où on force la tuile (i, j) à avoir la tuile $(i+1, j)$ à sa droite et la tuile $(i, j + 1)$ au dessus. Il suffit alors de superposer les tuiles où $ij = 0$ aux tuiles de T et les autres aux tuiles de τ pour obtenir rapidement la NP-difficulté du problème.

Il est difficile d'estimer précisément la paternité de ces théorèmes. On trouve trace de PAV-FIN2 dès Levin [Lev73]. Une preuve de la NP-complétude de PAV-FIN3 est donnée dans Lewis [Lew78], et est mentionnée dans Garey-Johnson [GJ79]. Voir aussi [Grä89].

A noter que ces problèmes de pavabilité peuvent être vus comme les problèmes canoniques de la théorie de la NP-complétude, à l'instar de SAT. Cette approche est suggérée par van Emde Boas et Harel [vEB97, Har85], et appliquée par Lewis-Papadimitriou [LP98]. Levin [Lev86, Lev84] utilise la pavabilité comme premier exemple d'un problème NP-complet *en moyenne*. Plus récemment Borchert [Bor08] formalise entièrement la classe NP à partir d'un problème de pavabilité.

Il est à noter que ces résultats n'utilisent pas totalement toutes les propriétés du codage : Le codage proposé simule un calcul en temps t et en espace s en *exactement* $t \times s$. Une question naturelle survient alors : Ce résultat est-il optimal ? Peut-on mettre un calcul "plus grand" dans un carré correctement pavé de taille $t \times s$? Une caractérisation *précise* est possible [Mey05] : Il faut alors s'intéresser aux machines de Turing dont on borne le nombre de retours (*reversals*, nombre de fois où la tête change de direction).

Dans le cas fini, une particularité importante est que la taille de la plus petite zone avec un bord pavable correctement est exactement le temps mis par la machine de Turing pour s'arrêter. On a utilisé cette propriété pour caractériser exactement les périodes possibles d'un TSFT.

Définition 3.1

La période d'une configuration X est le plus petit entier n tel que X soit périodique de période n, i.e. $\mathsf{X} + (n, 0) = \mathsf{X} + (0, n) = \mathsf{X}$. Une configuration non périodique n'a pas de période.

On peut alors démontrer :

Théorème 3.7 (*[I4]*)

Soit $L \subseteq \mathbb{N}^\star$ et $un(L) = \{a^n | n \in L\}$.

Alors L est l'ensemble des périodes d'un TSFT si et seulement si $un(L) \in \mathrm{NP}$. Plus exactement, si $un(L) \in \mathrm{NTIME}(n^d)$ alors il existe un TSFT en dimension $2d$ dont L est l'ensemble des périodes. Réciproquement, l'ensemble L des périodes d'un TSFT en dimension d vérifie $un(L) \in \mathrm{NTIME}(n^d)$.

La remarque précédente sur les "reversals" explique pourquoi il est impossible d'obtenir un théorème plus précis. A contrario, si L est l'ensemble des périodes d'un espace de pavages sofique, $un(L)$ peut être n'importe quel ensemble récursivement énumérable. La construction de [I4] est complexe et nécessite plusieurs outils, en particulier un jeu de tuiles apériodique déterministe, dont nous reparlerons à la fin de ce chapitre. On peut cependant utiliser cette méthode pour donner une nouvelle preuve de l'indécidabilité de la pavabilité du plan de façon périodique, preuve présentée dans [A1], et découverte également par Kari.

A noter que ce dernier théorème est le seul théorème de cette partie à parler précisément d'espaces de pavages (plus exactement de TSFT). Tous les autres problèmes de complexité et d'indécidabilité concernent uniquement des variantes du problème de pavabilité, mais jamais les espaces de pavages en eux-mêmes. Ce codage est en effet inadapté. Par exemple, le jeu de tuiles de la figure 3.1 produit des tas de pavages sans calcul (sans tête), ou au contraire avec plusieurs têtes. Les seuls pavages que nous contrôlons précisément sont uniquement ceux qui contiennent une certaine tuile au centre. Le but des constructions que nous allons voir à présent est donc de controler très précisément les constructions sans calcul Turing.

3.2 Codage auto-similaire

Les constructions présentées ici ont été introduites dans le but de montrer l'indécidabilité de la pavabilité du plan. On cherche donc à encoder une machine de Turing M dans un jeu de tuiles de Wang τ_M, de sorte que τ_M pave le plan si et seulement si M ne s'arrête pas. Intuitivement, il est donc nécessaire que tous les pavages par τ_M (s'ils existent) contiennent un calcul Turing. Si on cherche à encoder une machine de Turing comme précédemment, il est nécessaire d'encoder *plusieurs* calculs dans un même pavage : S'il existe des zones arbitrairement grandes du pavage sans calcul (ie sans tête de machine de Turing), il existera un pavage sans calcul. Toutes les constructions de ce type utilisent donc un mécanisme pour régler ce problème. En général, il s'agit de constructions *autosimilaires*. On ne cherchera pas à définir cette expression ambiguë, et on se contentera d'exemples.

La première construction est dûe à Berger [Ber64], mais de nombreuses autres constructions ont été proposées, souvent avec les mêmes propriétés [Rob71, Oll08, DSR08]. Il ne s'agit pas ici de détailler toutes les méthodes et nous nous focaliserons d'abord sur la preuve de Robinson [Rob71].

De nombreux articles ont été consacrés à cette construction [JM97, Sal89, AD01] et on ne va pas les répéter, on s'intéressera uniquement au dessin de la figure 3.4. Tout pavage par le jeu de tuiles de Robinson produit des carrés imbriqués les uns dans les autres. L'idée est de coder un calcul dans chacun des carrés (rouges sur le dessin) finis, et ce en évitant les carrés plus petits qu'il contient. On peut en effet démontrer que tout carré rouge de taille $4^n + 1$ contient une grille (n'intersectant pas les carrés rouges plus petits) d'une taille d'exactement 2^n.

On peut donc coder 2^n étapes du calcul d'une machine de Turing dans un carré de taille 4^n. Si on cherche à savoir si cette machine de Turing s'arrête, on obtient donc :

Théorème 3.8 (*[Ber64]*)

Le problème suivant est indécidable :

NOM :	PAV
ENTRÉE :	Un jeu de tuiles de Wang τ
SORTIE :	Est-ce qu'il existe un pavage du plan par τ ?

A noter que dans cette construction, les cellules qui contiendront du calcul sont exactement celles qui sont dans un carré de taille 4^n et qui ne sont contenues dans aucun carré de taille plus petite. En particulier, le jeu de tuiles de Robinson peut produire un pavage qui contient des carrés infinis ou, plus exactement, deux demi droites formant le bord d'un carré infini. S'il ne s'agit pas du bord bas et gauche qui définit le début du calcul de la machine de Turing, alors ces cellules peuvent contenir absolument n'importe quoi. Ainsi, dans cette construction, il existe des pavages qui contiennent "plus" que le calcul de la machine de Turing. En particulier, le jeu

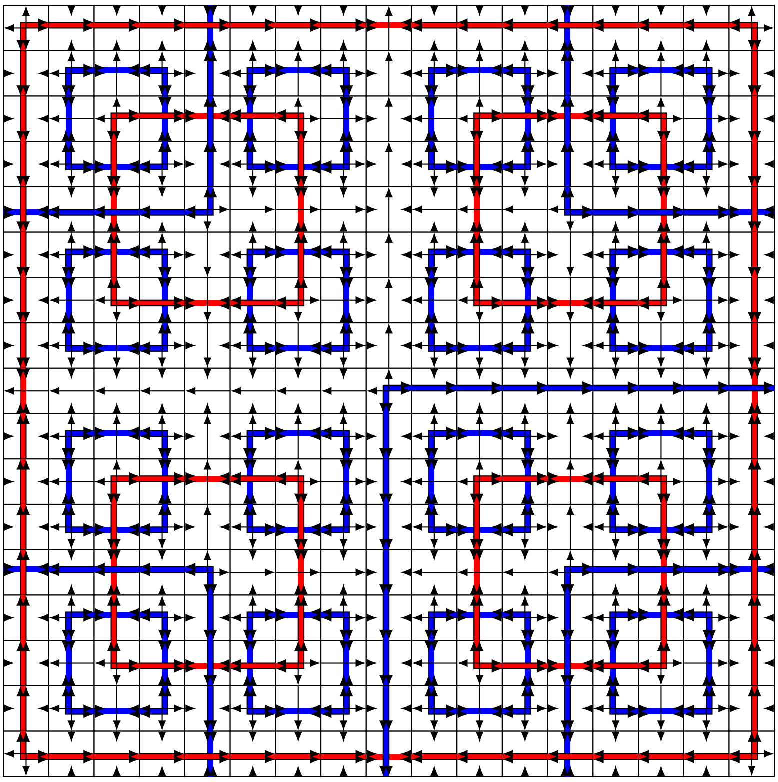


FIGURE 3.4 – Morceau d'un pavage de Robinson

de tuiles obtenu ne contiendra pas que des configurations quasipériodiques. Bien que la preuve d'indécidabilité soit bien claire, certains des pavages produits peuvent donc être complexes, et donc non maîtrisés.

Problème 3

> Peut-on encoder une machine de Turing M en un jeu de tuiles de Wang τ_M de sorte que :
> – Si M s'arrête, τ_M ne pave pas
> – Si M ne s'arrête pas, τ_M pave le plan, et ce d'une seule manière (ie. l'espace de pavages engendré est minimal) ?

Contrairement au codage exact précédent, le codage obtenu par Robinson [Rob71] n'est pas un diagramme espace-temps infini, mais une infinité de diagrammes espace-temps fini. En particulier, si on cherche à coder une machine de Turing avec oracle, l'oracle peut a priori être différent entre chaque calcul. Si on veut représenter un seul et unique calcul, il faut *synchroniser* l'oracle entre les différents calculs. On trouvera dans [Mye74] une explication de comment procéder. On peut alors démontrer :

Théorème 3.9 (*[Mye74, Sim11b]*)

> Pour tout ensemble Π_1^0 S, il existe un jeu de tuiles de Wang τ tel que tout pavage par τ code un élément de S, et tout élément de S soit codé par un pavage de τ. En particulier si S ne contient aucun point récursif, τ ne contient aucun pavage récursif.

Il faut faire attention à bien comprendre en quoi ce théorème est plus faible que le théorème 3.2 dans le cas de la tuile fixée au centre. Dans notre cas ici, plusieurs pavages (une infinité non dénombrable !) représentent le même élément X de S. De plus les pavages peuvent être plus compliqués (au sens Turing) que l'élément X de S qu'ils représentent, mais au moins l'un d'entre eux est du même degré que X. En particulier, cette construction a la propriété que l'ensemble des degrés Turing obtenus par le jeu de tuiles est la clotûre vers le haut des degrés Turing de S, alors que dans le cas du théorème 3.2, on a *exactement* les mêmes degrés Turing que S.

3.3 Codage clairsemé

La construction présentée ici et dans l'annexe F a été conçue dans le but d'éviter d'avoir des pavages (calculatoirement) complexes qui codent des machines de Turing simples. Elle permet en particulier de répondre à la question suivante : Peut-on trouver un TSFT *dénombrable* avec du calcul ?

Elle possède deux propriétés
– Chaque calcul Turing est représenté par une seule configuration
– Les configurations sans calcul sont connues, et simples

L'idée de la construction est de représenter une grille *clairsemée*, c'est à dire que les cellules de la grille s'éloignent le plus en plus les unes des autres, comme illustré dans la figure 3.5. On peut donc clairement représenter un calcul Turing en utilisant les points d'intersection entre les lignes horizontales et verticales pour former une grille. Le TSFT clairsemé a la particularité que les autres configurations contiennent au plus une ligne horizontale et une ligne verticale. Autrement dit, elles ne contiennent pas de calcul.

La représentation de la figure 3.5 est symbolique. Pour réaliser la même construction avec un TSFT, il faut être assez prudent. Le TSFT est présenté partiellement sur la figure 3.6 : pour obtenir la construction, il faut faire un produit cartésien du TSFT sur la figure 3.6 avec lui-même

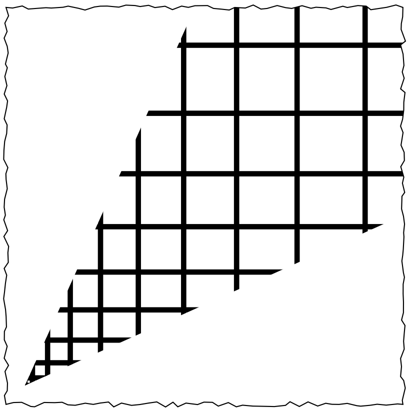


FIGURE 3.5 – Représentation symbolique d'une des configurations du TSFT clairsemé. Les autres configurations contiennent au plus une ligne horizontale et une ligne verticale.

tourné de 90 degrés, de sorte que les lignes horizontales et verticales ainsi obtenues forment la grille tant espérée. La construction est décrite plus précisément dans [I1].

Ce TSFT a la particularité que toutes ses configurations sont calculables ; en particulier il est dénombrable. On contrôle donc parfaitement ce qui arrive lorsqu'on y injecte du calcul :

Théorème 3.10 (*[I1]*)

Pour tout ensemble Π_1^0 S, il existe un jeu de tuiles de Wang τ tel que tout pavage par τ, sauf un ensemble dénombrable calculable, code un élément de S, et réciproquement tout élément de S est codé par un unique pavage par τ.

En particulier l'ensemble des degrés Turing des pavages par τ est égal à l'ensemble des degrés Turing des points de S, à l'ajout éventuel du degré des points calculables près.

En particulier, il existe des jeux de tuiles τ engendrant un nombre dénombrable de pavages et dont certains sont non calculables.

Ce théorème est à comparer au théorème 3.9 et au théorème 2.11 : L'ensemble des degrés Turing d'un jeu de tuiles peut être (quasi) quelconque s'il contient le degré des points récursifs, mais est fortement limité (il contient tout degré supérieur à $0'$) sinon.

Cette nouvelle construction est très récente, et on ne sait pas encore tout ce qu'elle peut permettre de démontrer. Lorsqu'il s'agit de montrer des propriétés de nature calculatoire sur les espaces de pavages, elle est cependant beaucoup plus manipulable que le codage autosimilaire, au sens où on maitrise totalement l'ensemble des pavages qu'elle engendre. Nous espérons en particulier l'utiliser afin de calculer la complexité exacte du problème de décider si un TSFT S est facteur d'un TSFT S'.

3.4 Autres

Pour finir, nous allons maintenant présenter deux codages un peu plus exotiques, où le diagramme espace-temps de la machine de Turing n'est pas directement visible.

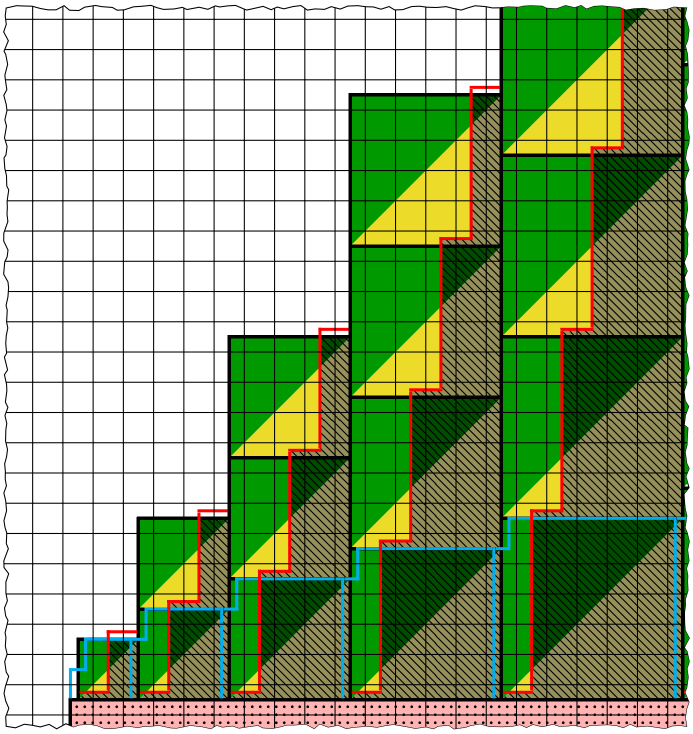


FIGURE 3.6 – La moitié du TSFT clairsemé. Les seules tuiles de Wang autorisées sont celles présentes sur ce dessin. Les verticales noires forment des segments de plus en plus espacés. Si on les croise avec des segments horizontaux construits de la même manière, on obtient la grille de la figure précédente. On obtient ainsi un jeu de 1522 (!) tuiles de Wang.

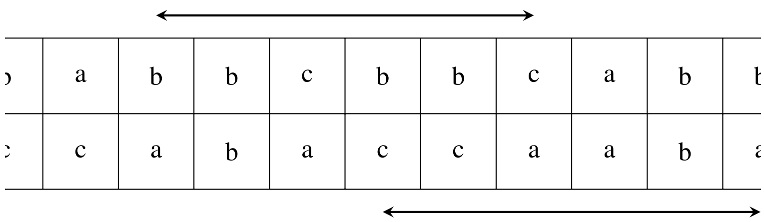


FIGURE 3.7 – Un $\Lambda_2^{(\theta)}$-système, comme introduit par Lewis [Lew79]. Il est donné par un alphabet Q et deux ensembles $A \subseteq Q^\theta$ et $B \subseteq Q^4$. Une paire de mots infinis x, y d'alphabets Q est valide si (a) tous les mots de longueur θ de x et y appartiennent à A et (b) si x_i, x_j, y_k, y_l sont 4 lettres à même distance sur chacun des rubans $(i - j = k - l)$ alors $(x_i, x_j, y_k, y_l) \in B$.

3.4.1 Double shift

Le premier d'entre eux a été introduit par Aanderaa et Lewis [AL74, Lew79] et est essentiellement unidimensionel. Avec des notations adaptées, on définit :

Définition 3.2

Soit Σ et Δ deux alphabets et S un sous-shift sofique sur l'alphabet $\Sigma \times \Delta$.
Le *double* shift défini par S est l'ensemble $\tilde{S}$ des $(x, y) \in \Sigma^{\mathbb{Z}} \times \Delta^{\mathbb{Z}}$ tel que pour toute translation $i \in \mathbb{Z}$, $(x, y + i) \in S$.

La terminologie double shift vient du fait que $\tilde{S}$ est clos par l'opération de translation usuelle, mais aussi par la translation d'une seule de ses deux composantes.

Le cas particulier utilisé par Lewis du double shift est décrit dans la figure 3.7.

On peut coder du calcul Turing dans un double shift :

Théorème 3.11 (*[AL74, Lew79]*)

Il est indécidable de savoir si un double shift est vide.

Si p est un nombre premier, notons $v_p(i)$ la valuation p-adique de i, c'est à dire la plus grande puissance de p qui divise i.

La preuve de Aanderaa et Lewis construit en particulier un double shift $\tilde{S}$ contenant un élément x, y pour lequel la cellule i de x (resp y) contient la cellule en coordonnées $(v_{11}(i), v_{13}(i))$ de la machine de Turing. L'idée est de considérer un alphabet $\{1, \dots, 10\}^2 \times Q \times \{1, \dots 12\}^2$. Les deux premières (resp deux dernières) composantes correspondent aux deux derniers chiffres non nuls de i en base 11 (resp. base 13) et la composante au milieu le symbole dans la cellule de la machine de Turing. Une fois l'objet compris, il est relativement simple de trouver les règles produisant le double shift indiqué, la difficulté principale étant dans la preuve que ces règles marchent comme prévu.

Ce codage est intéressant ici pour la raison suivante : Si $\tilde{S}$ est un double shift, alors l'ensemble des configurations bidimensionnelles où le symbole en position (i, j) est (x_i, y_{i+j}) (autrement dit x est identique sur chaque ligne, et y décalé vers la gauche à chaque étape) est un espace sofique, et on peut obtenir les règles définissant cet espace sofique en partant des règles définissant $\tilde{S}$, de sorte qu'on obtient encore une fois une preuve de l'indécidabilité de la pavabilité du plan.

Cette preuve est essentiellement différente des précédentes, puisque, si on regarde le codage obtenu, tout le calcul Turing est contenu dans *chaque* ligne du plan.

De plus, sans vouloir rentrer dans les détails, on remarque que dans le codage bidimensionel, les cellules sont divisées en deux parties, l'une constante sur chaque colonne, l'autre constante sur chaque antidiagonale. Si on regarde de près le codage sofique utilisé dans la preuve, on peut aussi s'apercevoir qu'on peut transformer le sofique en un TSFT en ajoutant une troisième composante, constante sur chaque ligne. Les TSFT obtenus sont donc constitués de trois parties, chacune constante dans une direction différente (horizontale, verticale, antidiagonale). En particulier, toute l'information de tout le pavage est entièrement contenue dans une diagonale, et le pavage obtenu peut donc être rendu nord-est déterministe, et permet ainsi de prouver, bien avant 1992 [Kar92], l'indécidabilité de la nilpotence des automates cellulaires, comme remarqué dans [AL74].

3.4.2 Immortalité

L'autre codage original est du à Kari [Kar07] et, bien qu'il code également un diagramme espace-temps, celui-ci n'est pas le diagramme usuel.

L'idée nouvelle, que Kari utilise déjà pour fabriquer son jeu de tuiles de Wang apériodique à 14 tuiles [Kar96, II96, ENP07], est de coder un système dynamique très simple :

Théorème 3.12

Soit $I = [a, b]$ un intervalle à bornes rationnelles, et f une fonction affine par morceaux de I dans I, où chaque coefficient impliqué est rationnel, et chaque morceau est borné par des points rationnels.

Alors il existe un TSFT S donné par tuiles de Wang qui code l'itération de f. Plus exactement, la ligne $i \in \mathbb{Z}$ code le point $f^i(x) \in I$ pour tout point $x \in I$.

L'idée de la preuve est la suivante : On code un réel x par son codage de Beatty : Si $x \in [0, 1]$, son codage est le mot infini $B(x)$ sur $\{0, 1\}^{\mathbb{Z}}$ défini par $B(x)_n = \lfloor nx \rfloor - \lfloor (n-1)x \rfloor$. On montre alors qu'on peut calculer le codage de Beatty de $f(x)$ à partir du codage de Beatty de x à partir d'un simple transducteur, qu'on transforme aisément en tuiles de Wang.

Kari obtient ainsi son jeu de tuiles apériodique à 14 tuiles en utilisant la fonction suivante : (le réel 1 a deux images par f, mais ça ne change rien au raisonnement)

$$f : \begin{array}{rcl} [\frac{1}{2}, 2] & \to & [\frac{1}{2}, 2] \\ x \leq 1 & \mapsto & 2x \\ x \geq 1 & \mapsto & \frac{2}{3}x \end{array}$$

On peut généraliser la construction à des fonctions de $\mathbb{R}^n$:

Théorème 3.13

Soit $I = \cup_n I_n = \cup_n \times_p [a_{n,p}, b_{n,p}]$ une union d'intervalles de $\mathbb{R}^p$ à bornes rationnelles, et f_n des fonctions affines de domaine I_n à coefficients rationnels. On note $f = \cup_n f_n$.

Alors il existe un TSFT S donné par tuiles de Wang, qu'on peut construire à partir de f, qui code l'itération de f. Plus exactement, la ligne $i \in \mathbb{Z}$ code le point $f^i(x) \in I$ pour tout point $x \in I$.

Or, il se trouve qu'il est indécidable de savoir, étant donné une telle fonction f affine par morceaux en dimension $p = 2$, s'il existe un point $x \in I$ tel que $f^i(x)$ est défini pour tout $i \in \mathbb{N}$ (autrement dit son orbite ne sort pas de I). On en déduit donc encore une fois l'indécidabilité du pavage du plan.

On encode en effet facilement une machine de Turing d'ensemble d'états Q et avec k symboles sur le ruban comme une fonction de domaine $Q \times [0,1] \times [0,1]$: la première partie correspond à son état, puis les deux suivantes, au contenu du ruban respectivement à gauche et à droite de la tête, encodé en base $K > k$. Le déplacement de la tête correspond à multiplier ou diviser par K. On peut ensuite voir cette fonction comme une fonction affine par morceaux de domaine $[|Q|, |Q|+1] \times [0,1]$.

Il faut maintenant bien faire attention qu'on ne peut pas, dans ce modèle, choisir le point x de départ de l'itération f : On cherche à savoir s'il existe un point x pour lequel $f^i(x)$ est défini pour tout x, et non pas si $f^i(x)$ est défini pour un point x spécifique. En particulier, on ne se demande pas si la machine de Turing, partant de la configuration initiale, ne s'arrête pas, mais s'il *existe* une configuration à partir de laquelle la machine de Turing ne s'arrête pas. Ce problème s'appelle le problème de la(l') (im)mortalité et a été montré indécidable par Hooper [Hoo66], autre étudiant de Hao Wang. La preuve implique en particulier l'existence des machines de Turing dont aucune trajectoire infinie n'est périodique, voir aussi [BCN02].

Du point de vue des pavages, cette construction est différente, au sens où maintenant chaque ligne correspond à une configuration, et le pavage correspond donc à un unique calcul d'une machine de Turing, mais partant d'une configuration arbitraire.

Il est à noter que cette construction, bien qu'élégante, n'est pas encore totalement comprise : On ne sait pas si le jeu de 14 tuiles exhibé par Kari ne produit par exemple que des configurations quasipériodiques. En terme de machines de Turing, remarquons que l'ensemble des configurations sur lesquelles une machine de Turing ne s'arrête pas est un ensemble Π^0_1. Que peut-on dire de plus ?

QUATRIÈME CHAPITRE

LOGIQUE

Dans cette partie, nous allons donner et expliquer certains des liens qui unissent la logique (et la théorie des modèles) à la théorie des pavages.

Comme il l'a été précisé, l'introduction du modèle des pavages par Hao Wang [Wan61] cherche à résoudre des problèmes de décision de certaines formules logiques par cet outil. La problématique se situe donc ainsi : A toute formule ϕ de type $\forall\exists\forall$, on peut associer un jeu de tuiles τ tel que ϕ est vraie si et seulement si τ pave le plan (en un sens à préciser). Ce problème de pavages se codant par une formule de type $\forall\exists\forall$, on obtient en particulier l'équivalence entre décider la satisfaction de ce type de formules, et décider la pavabilité (pour une forme très spécifique de pavabilité). Nous étudierons cette connexion dans l'introduction.

Dans la suite, on s'intéresse alors à considérer les pavages (resp. les espaces de pavages) comme des structures (resp. des théories) et on essaiera d'examiner comment la plupart des problématiques de pavages se traduisent naturellement dans un contexte logique.

Un peu d'histoire

Profitons de cette partie spécifiquement décidée à la logique pour expliquer historiquement pourquoi les pavages ont été introduits.

Considérons une relation binaire S et une relation ternaire T et considérons la formule suivante :

$$\phi = \forall x, \exists y, \forall z, (S(x,y) \wedge T(x,y,z) \wedge \neg T(y,y,z)) \vee \neg S(x,y)$$

On se demande s'il existe un modèle $\mathfrak{M}$ où la formule ϕ est vraie. Il faut donc pour cela choisir un ensemble M, définir les relations $S^{\mathfrak{M}} \subseteq M \times M$ et $T^{\mathfrak{M}} \subseteq M \times M \times M$ et vérifier que $\mathfrak{M}$ vérifie ϕ.

Pour analyser la formule, considérons un modèle $\mathfrak{M}$ où ϕ est vraie. On peut donc trouver, pour chaque x, un y qui rend vraie la formule. Notons, par abus, $f(x)$ cet élément y, ce qui revient à faire une skolemisation.

Partons alors d'un élément x_0 et considérons l'ensemble $A = \{f^n(x_0), n \in \mathbb{N}\}$. On s'aperçoit alors naturellement que cette sous-structure $\mathfrak{A}$ de $\mathfrak{M}$ vérifie également la formule ϕ.

Considérons maintenant la structure sur $\mathbb{N}$ définie de la façon suivante :

- $(n,m) \in S^{\mathbb{N}}$ ssi $(f^n(x_0), f^m(x_0)) \in S^{\mathfrak{M}}$
- $(n,m,p) \in T^{\mathbb{N}}$ ssi $(f^n(x_0), f^m(x_0), f^p(x_0)) \in T^{\mathfrak{M}}$

Autrement dit, on identifie n et $f^n(x_0)$. Il est alors clair que $\mathbb{N}$ vu ainsi vérifie la formule ϕ et vérifie plus exactement :

$$\psi = \forall x, \forall z, (S(x,x+1) \wedge T(x,x+1,z) \wedge \neg T(x+1,x+1,z)) \vee \neg S(x,x+1)$$

Ainsi, savoir s'il existe un modèle où la formule ϕ est vraie revient à savoir si on peut définir S et T sur $\mathbb{N}$ de sorte que ψ soit vraie. Une fois qu'on a obtenu cette formule ψ, il n'est maintenant plus très difficile de convertir la satisfiabilité de ψ en un problème de pavabilité.

A x et z donné, la véracité de la formule dépend uniquement de la valeur de T et de S sur les valeurs $x, x+1$ et z.

Pour chaque x et z notons $C(x,z)$ la liste des valeurs de vérités de T et S en tous ces points, c'est à dire la valeur de $S(a,b)$ et $T(a,b,c)$ pour $(a,b,c) \in \{x, x+1, z\}^3$. $C(x,z)$ ne peut donc prendre qu'un nombre *fini* de valeurs différentes.

Ces termes sont soumis aux contraintes suivantes :

- Les valeurs de vérité $C(x,z)$ doivent être choisies de sorte que la formule ψ soit vraie en (x,z).
- $C(x,z)$ doit être en accord avec $C(x+1,z)$ sur toutes les valeurs qu'elles ont en commun
- $C(x,z)$ doit être en accord avec $C(x,z+1)$ sur toutes les valeurs qu'elles ont en commun : La valeur de $T(x,x+1,x+1)$ des 2 termes doit par exemple ainsi coincider.
- Si $z = x$ ou $z = x+1$, alors les valeurs de vérité de $C(x,z)$ en tiennent compte : Si $z = x$, la valeur de vérité de $T(x,x,x+1)$ doit être la même que celle de $T(x,z,x+1)$

Supposons maintenant données pour chaque $C(x,z)$ avec $z \geq x$ des valeurs de vérité vérifiant les contraintes précédentes. On voit alors qu'elles permettent de définir de façon non contradictoire une valeur de vérité pour $T(\cdot,\cdot,\cdot)$ et $S(\cdot,\cdot)$ en tout point prouvant en fait que la formule ψ est satisfaisable.

De plus, ces contraintes sont *locales*, puisque la valeur de vérité en un point ne dépend que des valeurs de vérité des points voisins, et du fait que ce point est sur la diagonale (les deux cas particuliers $z = x$ et $z = x+1$) ou non. Savoir si la formule ϕ est satisfaisable revient donc à résoudre un problème de *pavage d'un huitième de plan*, c'est à dire des cellules (x,z) où $z \geq x$ par un nombre fini de couleurs (les $C(x,z)$) soumis à des contraintes locales et ayant des valeurs particulières sur la diagonale.

C'est ce raisonnement qui a conduit Hao Wang [Wan61] en 1961 à l'introduction des tuiles qui portent désormais son nom. Le raisonnement précédent peut être fait pour toute formule de la forme $\forall\exists\forall$: La satisfaction d'une telle formule revient à la pavabilité d'un huitième de plan, la diagonale étant soumis à des contraintes particulières. En fait Wang procède dans son article différemment de cette exposition : il aboutit au problème du pavage d'un quart de plan, la première colonne ayant des contraintes particulières.

Wang s'intéresse alors à la pavabilité du plan, qui est bien plus naturelle que cette forme batarde de pavabilité. Il pose alors sa fameuse conjecture :

> 4.1.2 *The fundamental conjecture*: A finite set of plates is solvable (has at least one solution) if and only if there exists a cyclic rectangle of the plates; or, in other words, a finite set of plates is solvable if and only if it has at least one periodic solution.

Il est intéressant de noter que les deux problèmes de pavabilité sont différents. En particulier, il n'est pas clair que l'existence d'un algorithme de décision pour l'un de ces problèmes entraîne l'existence d'un algorithme pour l'autre : On pourrait pouvoir décider la pavabilité du plan sans pouvoir décider la pavabilité d'un quart de plan avec une contrainte sur la première colonne. Wang ne s'aperçoit que peu de cette distinction, et la traite d'un simple “Some change is needed to take care of the additional conditions on the first two columns”.

En ce qui concerne le fragment logique, Büchi [Büc62a] montre l'indécidabilité de $\forall\exists\forall \wedge \exists$ par réduction aux machines de Turing. On pourrait l'obtenir de nos jours à partir de l'indécidabilité du pavage du plan contenant une tuile particulière : Les quantificateurs $\forall\exists\forall$ décrivant le jeu de tuiles, le quantificateur $\exists$ obligeant la tuile particulière à apparaître. Büchi n'arrive cependant pas à se passer de ce dernier quantificateur, mais s'aperçoit qu'une formule $\forall\exists\forall$ peut tout de même forcer l'existence de certaines tuiles :

> It is important to realize that also in conditions of form $\forall x, y\underline{M}(x, x', y)$ one still has use of one axis, namely one can formulate special restraints on the diagonal! In September 61 the author explained the situation to HAO WANG. He now claims in collaboration with A.S.KAHR AND E.F MOORE, to have found a construction $M \to \underline{M}$ as required in the above problem.

Quelques mois plus tard en effet, Kahr, Moore et Wang [KMW62] montrent que la pavabilité d'un huitième de plan (avec une diagonale particulière) est indécidable, puis que la pavabilité d'un huitième de plan s'exprime par une formule $\forall\exists\forall$, de sorte que la satisfaction d'une formule $\forall\exists\forall$ soit indécidable.

> LEMMA 2. If the diagonal-constrained domino problem is unsolvable, then the decision problem of the restricted $\forall\exists\forall$ case is unsolvable.

> LEMMA 3. The unsolvability of the diagonal-constrained domino problem: we give a general procedure by which, given any Turing machine Z and a fixed-initial state q_1, we can find a pair (P, Q) of domino sets having the mirror property such that, beginning with a blank tape in the state q_1, Z halts eventually if and only if P has no diagonal-constrained solution in half of the first quadrant relative to Q.

Le cas de $\forall\exists\forall$ est clos. Cependant le problème de la pavabilité du plan ne l'est pas encore. Berger, alors en thèse avec Wang, montrera en 1964 (publié en 1966) [Ber64] que le problème de la pavabilité du plan est également indécidable. Quelques années plus tard, Aanderaa et Lewis[AL74, Lew79], s'intéressant toujours à ce même fragment logique $\forall\exists\forall$, apportent une nouvelle démonstration de l'indécidabilité du pavage de plan, peu connue à ce jour.

4.1 Structures

Il y a principalement deux façons de représenter des configurations comme des structures : on peut les voir comme des structures sur $\mathbb{Z}$ ou sur $\mathbb{Z}^2$. Dans ce chapitre, on les verra comme structures sur $\mathbb{Z}^2$.

La signature τ considérée est constituée

- de 4 fonctions unaires, représentant les déplacements dans les 4 directions et notées astucieusement $\mathsf{North}, \mathsf{South}, \mathsf{East}, \mathsf{West}$;
- de prédicats unaires P_c pour chaque couleur c, $P_c(x)$ signifiant que x est de couleur c.

On associe naturellement à chaque configuration M une structure $\mathfrak{M}$ en interprétant correctement les différents éléments de la signature. On peut ainsi vérifier que les formules suivantes sont vraies dans la structure $\mathfrak{M}$ correspondant à la configuration M de la figure 4.1 :

$$\exists x, P_{\textcolor{cyan}{\blacksquare}}(x) \wedge P_{\textcolor{red}{\blacksquare}}(\mathsf{North}(x))$$

$$\forall x, \mathsf{North}(\mathsf{South}(x)) = x$$

Les pavages sont définis par des propriétés *locales*. Cette localité colle tout-à-fait avec le caractère local de la logique du premier ordre, localité qui s'exprime à travers par exemple le lemme de Hanf [I8, Han65, EF95, Lib04] ou le théorème de Gaifman. Elle se traduit dans les

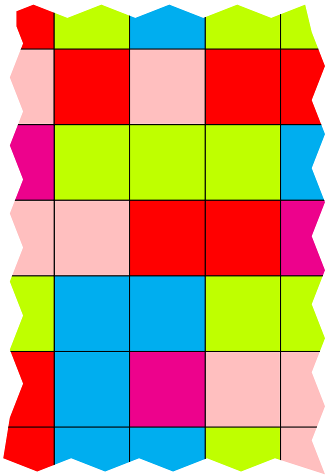


FIGURE 4.1 – Une configuration M

pavages par le résultat suivant :

Définition 4.1

Soit n et k deux entiers. Deux configurations M et N sont dites (n, k)-équivalentes si pour tout motif P de taille n :

- Si P apparaît dans M strictement moins de k fois, alors il apparaît exactement le même nombre de fois dans M et N
- Si P apparaît dans M plus de k fois, alors il apparaît plus de k fois dans N

Cette relation est bien une relation d'équivalence. A n et k fixés, elle a un nombre fini de classes.

Le lemme de Hanf peut alors s'écrire :

Lemme 4.1

Pour toute formule ϕ du premier ordre, il existe (n, k) tel que

$$\text{si M et N sont } (n, k)\text{-équivalents, alors}$$

$$\mathfrak{M} \models \phi \iff \mathfrak{N} \models \phi$$

Autrement dit, l'ensemble des configurations modèles d'une formule du premier ordre est une union (finie) de classe d'équivalences. Dit autrement : *toute formule est équivalente, sur les configurations, à une combinaison booléenne de formules du type $\phi_{\geq k}(P)$ et $\phi_{\leq k}(P)$, spécifiant que le motif P apparaît plus de (resp. moins de) k fois dans la configuration.*

C'est en particulier cette remarque qui explique en quoi les pavages se prêtent naturellement à une étude logique : La théorie des pavages, comme la logique du premier ordre, ne peut parler que de propriétés *locales* et donc d'apparition (ou non) de motifs.

Par exemple, deux configurations avec les mêmes motifs sont dans la même classe d'équivalence, ce pour tout n et k.

En particulier

Corollaire 4.2

> M et N ont les mêmes motifs (sont *localement isomorphes*) si et seulement si $\mathfrak{M}$ et $\mathfrak{N}$ sont élémentaires équivalentes (i.e. ont la même théorie).

Cela ne signifie pas que les seules structures élémentairement équivalentes à un modèle $\mathfrak{M}$ sont uniquement celles correspondant aux configurations localement isomorphes. Ceci est faux ne serait-ce que qu'en vertu du théorème de Lowenheim-Skolem ascendant : il existe des structures ayant la même théorie que $\mathfrak{M}$ de cardinal infini quelconque, donc en particulier qui ne représentent pas des configurations).

4.2 Espaces de pavages

Quand on se donne une théorie T, sous quelles hypothèses est-ce que l'ensemble des modèles de T (plus exactement l'ensemble des configurations modèles de T) est-il un espace de pavages ?

L'idée est simple :

Définition 4.2

> Si T est un ensemble de formules, on note $S(T)$ l'ensemble des configurations vérifiant T

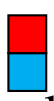
A noter que T, en tant que théorie, n'a pas que des configurations comme modèles ; en particulier si T est universelle, l'union disjointe de deux configurations modèles de T est modèle de T.

Théorème 4.3

> Si T est un ensemble de formules universelles (de la forme $\forall x_1 \dots x_n \phi(x_1 \dots x_n)$ avec ϕ sans quantificateurs), alors $S(T)$ est un espace de pavages. Réciproquement, pour tout espace de pavages S, il existe un ensemble T tel que $S(T) = S$.

Plus exactement, on peut démontrer que toute formule universelle est équivalente à une combinaison (positive) de formules de la forme $\phi_{\leq k}(P)$ [I6], et une telle formule représente bien un ensemble de pavages. Pour la réciproque, il suffit de remarquer que le fait que le motif P n'apparaît pas (ie est interdit) s'énonce en disant qu'en tout point x, le motif P n'est pas présent. Ainsi la formule

$$\forall x, \neg(P_{\blacksquare}(x) \wedge P_{\blacksquare}(\mathsf{North}(x)))$$

signifie que le motif  n'apparaît pas.

A remarquer que les espaces de pavages correspondant à un nombre *fini* de formules ne correspondent pas aux TSFT, puisque par exemple on peut exprimer avec une seule formule le fait qu'au plus un point rouge apparaît, ce qui correspond à un espace sofique qui n'est pas de type fini. En revanche il est clair que les espaces de pavages effectifs correspondent aux ensembles de formules récursivement énumérables.

L'intérêt d'une approche logique est de pouvoir caractériser certaines classes de pavages par des fragments logiques. On peut ainsi démontrer :

Théorème 4.4 (*[I6]*)

Si T est une formule monadique de la forme $\exists X_1 \dots \exists X_n \forall x_1 \dots \forall x_p \phi(X_1 \dots X_n, x_1 \dots x_p)$, alors l'espace $S(T)$ est sofique. Réciproquement, si S est sofique, alors il existe T de la forme précédente, avec un seul quantificateur au premier ordre, tel que $S(T) = S$.

Pour montrer ce théorème, on prouve une variante du lemme de Hanf adapté à ce type de formule.

Réciproquement, regarder des fragments logiques particuliers permet de découvrir de nouvelles classes intéressantes d'espace de pavages :

Définition 4.3

- Un espace de pavages S est à seuil s'il existe un nombre fini de paires (M_i, p_i) de motifs et d'entiers de sorte que S est exactement l'ensemble des configurations contenant moins de p_i occurences du motif M_i. Les unions d'espaces à seuil sont précisément les espaces reconnus au premier ordre universel.
- – Un espace de pavages est Π_0 (resp. Σ_0) si c'est un TSFT
  – Un espace de pavages S est Π_k (resp Σ_k) s'il existe un espace Σ_{k-1} (resp. Π_{k-1}) $\tilde{S}$ sur $(\Sigma \times \Delta)^{\mathbb{Z}^2}$ tel que X est dans S si et seulement si pour tout (resp. il existe) Y dans $\Delta^{\mathbb{Z}^2}$, $\mathsf{X} \times \mathsf{Y} \in \tilde{S}$.
  – Les Π_1 sont les TSFT.
  – Les Σ_1 (et donc Σ_2) sont les sofiques
- Si S est un TSFT et $\mathsf{M}_1, \mathsf{M}_2$ deux motifs distincts, le TSFT doublement pointé $(S, \mathsf{M}_1, \mathsf{M}_2)$ est l'ensemble des configurations de S contenant au moins une fois le motif M_1 et au moins une fois le motif M_2. Un ensemble de configurations est reconnu au second ordre monadique si et seulement si c'est la projection d'un TSFT doublement pointé.

Les théorèmes donnant lieu à ces définitions se trouve dans l'annexe H.

Une question intéressante arrive ainsi naturellement, concernant les espaces de pavages Π_2 : On ne connaît aucun exemple d'espace de pavages de cette classe qui ne soit pas un sofique, et il faudrait comprendre ces espaces. En particulier, est-ce que la hiérarchie des Π_k/Σ_k s'effondre ?

Il est assez étonnant que la classe des pavages doublement pointés n'ait jamais été étudiée. Dès leur origine, on étudie ainsi les pavages simplement pointés, c'est à dire les pavages par tuiles de Wang avec tuile spécifiée au centre (voir en particulier le théorème 3.1 du chapitre précédent), mais jamais cette variante. A noter que les pavages pointés par plus de deux points ne sont pas intéressants, puisqu'ils peuvent se coder en pointant exactement deux points [I6]. Cette classe a des propriétés remarquables : décider si un ensemble pointé est vide est Σ_2-complet pour la hiérarchie arithmétique, et on peut même montrer que tout ensemble Σ_2^0 de $\{0,1\}^{\mathbb{N}}$ est isomorphe à l'ensemble des configurations d'un ensemble pointé (on généralise ainsi le théorème 3.2 du chapitre précédent). On trouve mention d'une classe similaire, mais avec des propriétés algébriques moins intéressantes dans Harel [Har85, Har86], où on s'intéresse aux pavages du quart de plan où la première ligne est de la forme $c_0^n c_1^\omega$.

Comme on peut le voir, l'approche logique, au lieu de donner des solutions, donne surtout beaucoup de nouveaux problèmes à étudier !

4.3 Théories et modèles

Pour finir, on va chercher ici à se demander comment les résultats d'indécidabilité sur les pavages peuvent être utilisés en logique.

Remarquons d'abord, encore une fois que l'ensemble des modèles d'une théorie T et l'ensemble des configurations vérifiant T sont des ensembles très différents. D'abord parce qu'il faut que T spécifie que nord est l'inverse de sud, que chaque cellule contient exactement une seule couleur, et qu'aucun choix non trivial de directions obtenu par composition de nord, sud, est, ouest ne permet de revenir à son état initial. Cependant, toutes ces contraintes peuvent s'exprimer facilement au premier ordre. En revanche, il est impossible d'exprimer le fait que deux cellules x, y peuvent toujours être reliées par composition bien choisie de nord, sud, est, ouest. Ainsi, l'union de deux configurations vérifiant une théorie T correspondant à des pavages vérifiera en général aussi T.

Soit maintenant C un ensemble de couleurs fixé dans toute la suite. Commençons donc d'abord par définir la théorie T_u, constituée d'un nombre fini d'axiomes, exprimant le fait que nord est l'inverse de sud et que chaque cellule contient une et une seule couleur. On remarque assez facilement que toutes les formules de T_u sont, dans notre théorie, des formules universelles. Si on remplace les fonctions par des prédicats binaires, elles deviennent des formules du type $\forall\exists\forall$.

Les modèles de T_u *connexes* (où tous les points sont reliés par application de nord, sud, est, ouest) sont de deux types : des configurations, ou des configurations "repliées" : Par exemple, si $\mathsf{North}^{15}(x) = x$, le modèle correspond à une bande de hauteur 15 qui se replie sur elle-même (un cylindre). On peut toutefois associer à ces configurations spéciales la configuration "dépliée". Comme tout modèle est une union de structures connexes, on peut donc toujours voir un modèle de T_u comme une union de configurations, repliées ou non.

Pour chaque TSFT S sur C, on peut lui associer une théorie universelle T_S exprimant que les motifs interdits par S n'apparaissent pas. Si la signature est choisie relationnelle plutôt que fonctionnelle, la théorie sera du type $\forall^k x$, où k est le diamètre du SFT (distance maximum entre deux points d'un motif interdit).

On remarque alors les choses suivantes [I8] :

– Si S pave le plan, $T_S \cup T_u$ a un modèle
– Si S ne pave pas le plan, $T_S \cup T_u$ n'a pas de modèles : En effet tout modèle de T_u peut se voir comme une ou plusieurs configurations, éventuellement repliées, et chacune d'entre elles vérifie T_S (vus les axiomes de T_S)
– Si S pave le plan de façon périodique, T_S a un modèle fini, et réciproquement (clair).

On a donc démontré :

Théorème 4.5

> On peut associer à tout TSFT S une théorie T_S de sorte que T_S est satisfiable si et seulement si S pave le plan.
>
> Si S est donné par tuiles de Wang, on peut supposer que T_S est une théorie relationnelle constituée de formules du type $\forall\exists\forall$.

Ce théorème montre ainsi qu'il est indécidable de savoir si une formule du type $\forall\exists\forall$ a un modèle [BGG01],

C'est un des premiers exemples d'utilisation des pavages pour montrer qu'un fragment logique (au premier ordre, ou pour des théories plus exotiques, voir [Grä89, vEB97]) est indécidable.

Notons que la preuve usuelle de l'indécidabilité du pavage du plan code une machine de Turing M en un jeu de tuiles de Wang τ_M de sorte que M ne termine pas si et seulement si τ_M

pave le plan. On code ici τ_M en une théorie T_M de sorte que τ_M pave le plan si et seulement si T_M a un modèle Si jamais on réussissait à transformer τ_M de sorte qu'il ne produise qu'un seul pavage quasipériodique (voir la discussion suivant le théorème 3.8), on pourrait s'arranger pour que, de plus, T_M soit une théorie complète si τ_M pave. Cependant ce résultat n'est pas encore, à ma connaissance, acquis.

Notons en particulier que les cas où T_S est une théorie complète sont bien compris : T_S est une théorie complète si et seulement si S est minimal sans point périodique. Remarquons qu'une théorie complète finiment axiomatisable est décidable (chercher simultanément une preuve de ϕ et $\neg\phi$), donc en particulier si S est minimale, sa théorie est décidable, ce qui signifie que son langage est récursif.

Si T_S n'est pas complète, il est naturel de se demander à quoi ressemblent les théories complètes extensions de T_S. Il est à peu près clair qu'on peut associer à tout sous espace de pavages de S une théorie complète (la théorie de l'union de ses pavages), mais toutes les théories complètes ne peuvent pas s'obtenir comme ça : Si S est un jeu de tuiles qui crée des pavages où la couleur rouge apparaît une seule fois, la théorie disant “il existe deux points rouges” a des modèles (l'union de deux pavages avec une seule apparition de la couleur rouge) mais qui ne sont pas de la forme précédente. Il serait donc intéressant de savoir caractériser ces extensions.

Finissons par une (nouvelle) question intéressante : l'ensemble des configurations modèles d'une théorie T n'est pas toujours un espace de pavages. Prenons en revanche l'ensemble S des configurations tel que pour tout modèle $\mathfrak{M}$ de T, $\mathfrak{M} \cup \{S\}$ est un modèle de T. On peut montrer que c'est toujours un espace de pavages, si jamais T interdit les modèles repliés. Est-il de type fini ?

Annexes

Annexe A

INDEX

Annexe B

BIBLIOGRAPHIE

[AD01] C. Allauzen et B. Durand. *Tiling Problems*. Dans *The Classical Decision Problem* [BGG01], chapitre A, pages 407–420. 2001. ISBN 3540423249

[AJM09] M. Anselmo, N. Jonoska et M. Madonia. *Framed Versus Unframed Two-Dimensional Languages*. Dans *35th Conference on Current Trends in Theory and Practice of Computer Science, SOFSEM*, pages 79–92. 2009. doi:10.1007/978-3-540-95891-8_11

Résumé : *In this paper we study a class of factor-closed tiling recognizable languages (FREC) that corresponds to certain symbolic dynamical systems called sofic subshifts. This class of languages is a subclass of the two-dimensional class of tiling recognizable languages, denoted REC. Differently from REC, languages in FREC can be recognized without framing pictures with a special boundary symbol. We study (un)ambiguity properties of FREC in comparison to the ones in REC. We show that a frame surrounding each block provides additional memory that can enforce the size and the content of pictures and can change (un)ambiguity properties. Consequently, we propose several variations of unambiguity for languages in FREC which may be better suited to understand this class.*

[AL74] S. O. Aanderaa et H. R. Lewis. *Linear Sampling and the* $\forall\exists\forall$ *Case of the Decision Problem*. Journal of Symbolic Logic, tome 39(3), pages 519–548, 1974. See also [Lew79]. euclid.jsl/1183739186

[AS] N. Aubrun et M. Sablik. *Simulation of effective subshifts by two-dimensional SFT and a generalization*. Preprint.

[AS09] N. Aubrun et M. Sablik. *An order on sets of tilings corresponding to an order on languages*. Dans *26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009, February 26-28, 2009, Freiburg, Germany, Proceedings*, pages 99–110. 2009.

Résumé : *Traditionally a tiling is defined with a finite number of finite forbidden patterns. We can generalize this notion considering any set of patterns. Generalized tilings defined in this way can be studied with a dynamical point of view, leading to the notion of subshift. In this article we establish a correspondence between an order on subshifts based on dynamical transformations on them and an order on languages of forbidden patterns based on computability properties.*

[ATW03] J.-H. Altenbernd, W. Thomas et S. Wohrle. *Tiling systems over infinite pictures and their acceptance conditions*. Dans *Developments in Language Theory (DLT)*, *Lecture Notes in Computer Science*, tome 2450/2003, pages 127–146. Springer Berlin / Heidelberg, 2003.

Résumé : *Languages of infinite two-dimensional words (ω-pictures) are studied in the automata theoretic setting of tiling systems. We show that a hierarchy of acceptance conditions as known from the theory of ω-languages can be established also over pictures. Since the usual pumping arguments fail, new proof techniques* *are necessary. Finally, we show that (unlike the case of ω-languages) none of the considered acceptance conditions leads to a class of infinitary picture languages which is closed under complementation.*

[BCN02] V. D. Blondel, J. Cassaigne et C. Nichitiu. *On the presence of periodic configurations in Turing machines and in counter machines*. Theoretical Computer Science, tome 289(1), pages 573–590, 2002.

Résumé : *A configuration of a Turing machine is given by a tape content together with a particular state of the machine. Petr Kurka has conjectured that every Turing machine – when seen as a dynamical system on the space of its configurations – has at least one periodic orbit. In this paper, we provide an explicit counterexample to this conjecture. We also consider counter machines and prove that, in this case, the problem of determining if a given machine has a periodic orbit in configuration space is undecidable.*

[BDG88] J. Balcazar, J. Diaz et J. Gabarro. *Structural Complexity II*. Springer-Verlag. 1988.

[Ber64] R. Berger. *The Undecidability of the Domino Problem*. Thèse de doctorat, Harvard University. 1964.

[Ber66] R. Berger. *The Undecidability of the Domino Problem*. Numéro 66 dans Memoirs of the American Mathematical Society. The American Mathematical Society. 1966.

[Ber95] J. Berstel. *Axel thue's papers on repetitions in words: a translation*. Publications du LaCIM, tome 20.

[BGG01] E. Börger, E. Grädel et Y. Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer. 2001. ISBN 3540423249

[Bir12] M.-G. D. Birkhoff. *Quelques théorèmes sur le mouvement des systèmes dynamiques*. Bulletin de la SMF, tome 40, pages 305–323, 1912.

[Bor08] B. Borchert. *Formal Language characterizations of P, NP, and PSPACE*. Journal of Automata, Languages and Combinatorics, tome 13(3/4), pages 161–183, 2008.

Résumé : *Based on the notions of locality and recognizability for n-dimensional languages n-dimensionally colorable 1-dimensional languages are introduced. It is shown: A language L is in NP if and only if L is n-dimensionally colorable for some n. An analogous characterization in terms of deterministic n-dimensional colorability is obtained for P. The addition of one unbounded dimension for coloring leads to a characterization of PSPACE.*

[BP91] D. Beauquier et J.-E. Pin. *Languages and scanners*. Theoretical Computer Science, tome 84, pages 3–21, 1991. doi:10.1016/0304-3975(91)90258-4

Résumé : *On s'intéresse dans cet article à une nouvelle classe d'automates, baptisés scanners. Ces machines peuvent être considérées comme des modèles pour les calculs qui nécessitent seulement une information locale. On donne une caractérisation algébrique effective des langages reconnus par divers types de scanners et on étudie le lien avec la logique du successeur.*

[BT09] L. Boyer et G. Theyssier. *On local symmetries and universality in cellular automata*. Dans *STACS*, pages 195–206. 2009. doi:10.4230/LIPIcs.STACS.2009.1836

Résumé : *Cellular automata (CA) are dynamical systems defined by a finite local rule but they are studied for their global dynamics. They can exhibit a wide range of complex behaviours and a celebrated result is the existence of (intrinsically) universal CA, that is CA able to fully simulate any other CA. In this paper, we show that the asymptotic density of universal cellular automata is 1 in several families of CA defined by local symmetries. We extend results reviously established for captive cellular automata in two significant ways. First, our results* *apply to well-known families of CA (e.g. the family of outer-totalistic CA containing the Game of Life) and, second, we obtain such density results with both increasing number of states and increasing neighbourhood. Moreover, thanks to universality-preserving encodings, we show that the universality problem remains undecidable in some of those families.*

[BT10] L. Boyer et G. Theyssier. *On factor universality in symbolic spaces*. Dans *MFCS*, pages 209–220. 2010. doi:10.1007/978-3-642-15155-2_20

Résumé : *The study of factoring relations between subshifts or cellular automata is central in symbolic dynamics. Besides, a notion of intrinsic universality for cellular automata based on an operation of rescaling is receiving more and more attention in the literature. In this paper, we propose to study the factoring relation up to rescalings, and ask for the existence of universal objects for that simulation relation. In classical simulations of a system S by a system T, the simulation takes place on a specific subset of configurations of T depending on S (this is the case for intrinsic universality). Our setting, however, asks for every configurations of T to have a meaningful interpretation in S. Despite this strong requirement, we show that there exists a cellular automaton able to simulate any other in a large class containing arbitrarily complex ones. We also consider the case of subshifts and, using arguments from recursion theory, we give negative results about the existence of universal objects in some classes.*

[Büc62a] J. R. Büchi. *Turing-Machines and the Entscheidungsproblem*. Math. Annalen, tome 148(3), pages 201–213, 1962. doi:10.1007/BF01470748

[Büc62b] J. R. Büchi. *Weak Second Order Logic and Finite Automata*. Z. Math. Logik, Grundlag. Math., tome 6, pages 66–92, 1962. doi:10.1002/malq.19600060105

[CD04] J. Cervelle et B. Durand. *Tilings: recursivity and regularity*. Theor. Comput. Sci., tome 310(1-3), pages 469–477, 2004. doi:10.1016/S0304-3975(03)00242-1

Résumé : *We establish a first step towards a Rice theorem for tilings: for non-trivial sets, it is undecidable to know whether two different tile sets produce the same tilings of the plane. Then, we study quasiperiodicity functions associated with tilings. This function is a way to measure the regularity of tilings. We prove that, not only almost all recursive functions can be obtained as quasiperiodicity functions, but also, a function which overgrows any recursive function.*

[CDK08] D. Cenzer, A. Dashti et J. L. F. King. *Computable symbolic dynamics*. Mathematical Logic Quarterly, tome 54(5), pages 460–469, 2008. doi:10.1002/malq.200710066

Résumé : *We investigate computable subshifts and the connection with effective symbolic dynamics. It is shown that a decidable Π^0_1 class P is a subshift if and only if there exists a computable function F mapping $2^{\mathbb{N}}$ to $2^{\mathbb{N}}$ such that P is the set of itineraries of elements of $2^{\mathbb{N}}$. Π^0_1 subshifts are constructed in $2^{\mathbb{N}}$ and in $2^{\mathbb{Z}}$ which have no computable elements. We also consider the symbolic dynamics of maps on the unit interval.*

[CDTW10] D. Cenzer, A. Dashti, F. Toska et S. Wyman. *Computability of Countable Subshifts*. Dans *Computability in Europe (CiE)*, *Lecture Notes in Computer Science*, tome 6158, pages 88–97. 2010. doi:10.1007/978-3-642-13962-8_10

Résumé : *The computability of countable subshifts and their members is examined. Results include the following. Subshifts of Cantor-Bendixson rank one contain only eventually periodic elements. Any rank one subshift , in which every limit point is periodic, is decidable. Subshifts of rank two may contain members of arbitrary Turing degree. In contrast, effectively closed (Π^0_1) subshifts of rank two contain only computable elements, but Π^0_1 subshifts of rank three may contain members of arbitrary c. e. degree. There is no subshift of rank ω.*

[Cer02] J. Cervelle. *Complexité structurelle et algorithmique des pavages et des automates cellulaires*. Thèse de doctorat, Université de Provence. 2002.

Résumé : *Ce travail de thèse étudie la complexité des pavages et des automates cellulaires. L'analyse débute par des considérations structurelles : la quasipériodicité des pavages. A tout ensemble de tuiles qui pave le plan, on associe une fonction de quasipériodicité qui quantifie sa complexité. Tout d'abord, on montre que toute fonction raisonnable peut être capturée par un ensemble de tuiles et qu'il existe des pavages dont la fonction de quasipériodicité croît plus rapidement que n'importe quelle fonction récursive. Ensuite, on démontre un théorème de Rice pour les pavages : l'ensemble des ensembles de tuiles qui admettent les mêmes pavages qu'un autre fixé est indécidable et récursivement énumérable. Enfin, on transpose notre résultat dans le contexte des automates cellulaires. La seconde partie de notre travail concerne l'étude des automates cellulaires sous l'angle des systèmes dynamiques, et plus particulièrement des automates chaotiques. Les définitions usuelles classifiant les automates chaotiques ne sont pas satisfaisantes. Pour palier ce problème, on utilise deux nouvelles topologies. La première est dite de Besicovitch, et permet de supprimer la prédominance du motif central lors de l'étude de l'évolution de l'automate. On apporte plusieurs résultats, les premiers indiquant que notre nouvel espace de travail est acceptable à l'étude des automates cellulaires, en tant que systèmes dynamiques ; les seconds montrent que la notion de chaos subsiste, grâce à la définition de sensibilité aux conditions initiales, mais que les classes plus chaotiques sont vides. La seconde topologie employée est définie à l'aide de la complexité algorithmique. Le but de cette approche est d'avoir une distance qui traduit la facilité à calculer un élément à l'aide de l'autre. Nos résultats complètent les précédents. Ils attestent de manière formelle que les automates cellulaires ne peuvent pas modifier continûment l'information contenue dans une configuration, et surtout qu'ils sont incapables d'en créer.*

[Cha08] G. Chaitin. *The Halting Probability via Wang Tiles*. Fundamenta Informaticae, tome 86(4), pages 429–433, 2008.

Résumé : *Using work of Hao Wang, we exhibit a tiling characterization of the bits of the halting probability* Ω

[CM06] A. Carayol et A. Meyer. *Context-Sensitive Languages, Rational Graphs and Determinism*. Logical Methods in Computer Science, tome 2(2), pages 1–24, 2006. doi:10.2168/LMCS-2(2:6)2006

Résumé : *We investigate families of infinite automata for context-sensitive languages. An infinite automaton is an infinite labeled graph with two sets of initial and final vertices. Its language is the set of all words labelling a path from an initial vertex to a final vertex. In 2001, Morvan and Stirling proved that rational graphs accept the context-sensitive languages between rational sets of initial and final vertices. This result was later extended to sub-families of rational graphs defined by more restricted classes of transducers. languages. Our contribution is to provide syntactical and self-contained proofs of the above results, when earlier constructions relied on a non-trivial normal form of context-sensitive grammars defined by Penttonen in the 1970's. These new proof techniques enable us to summarize and refine these results by considering several sub-families defined by restrictions on the type of transducers, the degree of the graph or the size of the set of initial vertices.*

[CR98] D. Cenzer et J. Remmel. Π^0_1 *classes in mathematics*. Dans *Handbook of Recursive Mathematics - Volume 2: Recursive Algebra, Analysis and Combinatorics*, *Studies in Logic and the Foundations of Mathematics*, tome 139, chapitre 13, pages 623–821. Elsevier. 1998. doi:10.1016/S0049-237X(98)80046-3

[CR11] D. Cenzer et J. Remmel. *Effectively Closed Sets*. ASL Lecture Notes in Logic. 2011, in preparation.

[Das08] A. Dashti. *Effective Symbolic Dynamics*. Thèse de doctorat, University of Florida. 2008.

Résumé : *We investigate computable subshifts and the connection with effective symbolic dynamics. It is shown that a decidable Π_1^0 class P is a subshift if and only if there is a computable function F mapping $2^{\mathbb{N}}$ to $2^{\mathbb{N}}$ such that P is the set of itineraries of elements of $2^{\mathbb{N}}$. A Π_1^0 subshift is constructed which has no computable element. Moreover Π_1^0 subshifts with higher degrees of difficulty are constructed. We also consider the symbolic dynamics of maps on the unit interval.*

[Del11] M. Delacourt. *Rice's theorem for mu-limit sets of cellular automata*. Dans *ICALP*, pages 89–100. 2011. doi:10.1007/978-3-642-22012-8_6

Résumé : *Cellular automata are a parallel and synchronous computing model, made of infinitely many finite automata updating according to the same local rule. Riceś theorem states that any nontrivial property over computable functions is undecidable. It has been adapted by Kari to limit sets of cellular automata [7], that is the set of configurations that can be reached arbitrarily late. This paper proves a new Rice theorem for μ-limit sets, which are sets of configurations often reached arbitrarily late.*

[DKB04] J.-C. Delvenne, P. Kurka et V. D. Blondel. *Computational universality in symbolic dynamical systems*. Machines, Computations, and Universality 4th International Conference, MCU 2004, pages 104–115. doi:0.1007/978-3-540-31834-7_8

Résumé : *Many different definitions of computational universality for various types of systems have flourished since Turingś work. In this paper, we propose a general definition of universality that applies to arbitrary discrete time symbolic dynamical systems. For Turing machines and tag systems, our definition coincides with the usual notion of universality. It however yields a new definition for cellular automata and subshifts. Our definition is robust with respect to noise on the initial condition, which is a desirable feature for physical realizability. We derive necessary conditions for universality. For instance, a universal system must have a sensitive point and a proper subsystem. We conjecture that universal systems have an infinite number of subsystems. We also discuss the thesis that computation should occur at the edge of chaos and we exhibit a universal chaotic system.*

[DLP+10] D. Doty, J. H. Lutz, M. J. Patitz, S. M. Summers et D. Woods. *Intrinsic universality in self-assembly*. Dans *STACS*, pages 275–286. 2010. doi:10.4230/LIPIcs.STACS.2010.2461

Résumé : *We show that the Tile Assembly Model exhibits a strong notion of universality where the goal is to give a single tile assembly system that simulates the behavior of any other tile assembly system. We give a tile assembly system that is capable of simulating a very wide class of tile systems, including itself. Specifically, we give a tile set that simulates the assembly of any tile assembly system in a class of systems that we call* locally consistent*: each tile binds with exactly the strength needed to stay attached, and that there are no glue mismatches between tiles in any produced assembly. Our construction is reminiscent of the studies of* intrinsic universality *of cellular automata by Ollinger and others, in the sense that our simulation of a tile system T by a tile system U represents each tile in an assembly produced by T by a $c \times c$ block of tiles in U, where c is a constant depending on T but not on the size of the assembly T produces (which may in fact be infinite). Also, our construction improves on earlier simulations of tile* *assembly systems by other tile assembly systems (in particular, those of Soloveichik and Winfree, and of Demaine et al.) in that we simulate the actual process of self-assembly, not just the end result, as in Soloveichik and Winfree's construction, and we do not discriminate against infinite structures. Both previous results simulate only temperature 1 systems, whereas our construction simulates tile assembly systems operating at temperature 2.*

[DLS08] B. Durand, L. A. Levin et A. Shen. *Complex tilings.* Journal of Symbolic Logic, tome 73(2), pages 593–613, 2008. doi:10.2178/jsl/1208359062

Résumé : *We study the minimal complexity of tilings of a plane with a given tile set. We note that every tile set admits either no tiling or some tiling with $O(n)$ Kolmogorov complexity of its $(n \times n)$-squares. We construct tile sets for which this bound is tight: all $(n \times n)$-squares in all tilings have complexity at least n. This adds a quantitative angle to classical results on non-recursivity of tilings – that we also develop in terms of Turing degrees of unsolvability.*

[DPR+10] D. Doty, M. J. Patitz, D. Reishus, R. T. Schweller et S. M. Summers. *Strong fault-tolerance for self-assembly with fuzzy temperature*. Dans *FOCS*, pages 417–426. 2010. doi:10.1109/FOCS.2010.47

Résumé : *We consider the problem of fault-tolerance in nanoscale algorithmic self-assembly. We employ a standard variant of Winfreeś abstract Tile Assembly Model (aTAM), the two-handed aTAM, in which square tiles – a model of molecules constructed from DNA for the purpose of engineering self-assembled nanostructures – aggregate according to specific binding sites of varying strengths, and in which large aggregations of tiles may attach to each other, in contrast to the seeded aTAM, in which tiles aggregate one at a time to a single specially designated seed assembly. We focus on a major cause of errors in tile-based self-assembly: that of unintended growth due to weak strength-1 bonds, which if allowed to persist, may be stabilized by subsequent attachment of neighboring tiles in the sense that at least energy 2 is now required to break apart the resulting assembly, i.e., the errant assembly is stable at temperature 2. We study a common self-assembly benchmark problem, that of assembling an $n \times n$ square using $O(\log n)$ unique tile types, under the two-handed model of self-assembly. Our main result achieves a much stronger notion of fault-tolerance than those achieved previously. Arbitrary strength-1 growth is allowed, however, any assembly that grows sufficiently to become stable at temperature 2 is guaranteed to assemble into the correct final assembly of an $n \times n$ square. In other words, errors due to insufficient attachment, which is the cause of errors studied in earlier papers on fault-tolerance, are prevented absolutely in our main construction, rather than only with high probability and for sufficiently small structures, as in previous fault tolerance studies.*

[DRS10] B. Durand, A. Romashchenko et A. Shen. *Effective Closed Subshifts in 1D Can Be Implemented in 2D*. Dans *Fields of Logic and Computation*, numéro 6300 dans Lecture Notes in Computer Science, pages 208–226. Springer. 2010. doi:10.1007/978-3-642-15025-8_12

Résumé : *In this paper we use fixed point tilings to answer a question posed by Michael Hochman and show that every one-dimensional effectively closed subshift can be implemented by a local rule in two dimensions. The proof uses the fixed-point construction of an aperiodic tile set and its extensions.*

[DSR08] B. Durand, A. Shen et A. Romashchenko. *Fixed Point and Aperiodic Tilings* . Rapport technique TR08-030, ECCC. 2008.

Résumé : *An aperiodic tile set was first constructed by R. Berger while proving the undecidability of the domino problem. It turned out that aperiodic tile sets appear in many topics ranging from logic (the Entscheidungsproblem) to physics (quasicrystals) We present a new construction of an aperiodic tile set. The flexibility of this construction simplifies proofs of some known results and allows us to construct a "robust" aperiodic tile set that does not have periodic (or close to periodic) tilings even if we allow some (sparse enough) tiling errors. Our construction of an aperiodic self-similar tile set is based on Kleene's fixed-point construction instead of geometric arguments. This construction is similar to J. von Neumann self-reproducing automata ; similar ideas were also used by P. Gacs in the context of error-correcting computations.*

[Dur94] B. Durand. *Inversion of 2D cellular automata: some complexity results*. Theoretical Computer Science, tome 132(2), pages 387–401, 1994.

Résumé : *In this paper, we prove the co-NP-completeness of the following decision problem: Given a two-dimensional cellular automaton A (even with Von Neumann neighborhood), is A injective when restricted to finite configurations not greater than its length ? ? In order to prove this result, we introduce two decision problems concerning, respectively, Turing machines and tilings that we prove NP-complete. Then, we present a transformation of problems concerning tilings into problems concerning cellular automata.*

[Dur99] B. Durand. *Tilings and Quasiperiodicity*. Theoretical Computer Science, tome 221(1-2), pages 61–75, 1999. doi:10.1016/S0304-3975(99)00027-4

Résumé : *Quasiperiodic tilings are those tilings in which finite patterns appear regularly in the plane. This property is a generalization of the periodicity ; it was introduced for representing quasicrystals and it is also motivated by the study of quasiperiodic words. We prove that if a tile set can tile the plane, then it can tile the plane quasiperiodically – a surprising result that does not hold for periodicity. In order to compare the regularity of quasiperiodic tilings, we introduce and study a quasiperiodicity function and prove that it is bounded by* $x \mapsto x + c$ *if and only if the considered tiling is periodic. At last, we prove that if a tile set can be used to form a quasiperiodic tiling which is not periodic, then it can form an uncountable number of tilings.*

[EF95] H.-D. Ebbinghaus et J. Flum. *Finite Model Theory*. Springer Monographs in Mathematics. Springer, Berlin. 1995. ISBN 3540287876

[Elg61] C. C. Elgot. *Decision Problems of Finite Automata Design and Related Arithmetics*. Transactions of the American Mathematical Society, tome 98(1), pages 21–51, 1961. doi:10.1090/S0002-9947-1961-0139530-9

[ENP07] S. Eigen, J. Navarro et V. S. Prasad. *An aperiodic tiling using a dynamical system and Beatty sequences*. Dans *Recent Progress in Dynamics*, *MSRI Publications*, tome 54. Cambridge University Press. 2007.

Résumé : *Wang tiles are square unit tiles with colored edges. A finite set of Wang tiles is a valid tile set if the collection tiles the plane (using an unlimited number of copies of each tile), the only requirements being that adjacent tiles must have common edges with matching colors and each tile can be put in place only by translation. In 1995 Kari and Culik gave examples of tile sets with 14 and 13 Wang tiles respectively, which only tiled the plane aperiodically. Their tile sets were constructed using a piecewise multiplicative function of an interval. The fact the sets tile only aperiodically is derived from properties of the function.*

[Fag74] R. Fagin. *Generalized first-order spectra and polynomial-time recognizable sets*. Dans *Complexity of Computation* (R. Karp, réd.), *SIAM-AMS Proceedings*, tome 7, pages 43–73. 1974.

[Flo56] I. Flores. *Reflected Number Systems*. IRE Transactions on Electronic Computers, tome EC-5(2), pages 79–82, 1956.

Résumé : *Many papers have been written about the reflected binary system and it is well known in the computer field for analog-to-digital conversion. The method used in creating this system may be extended to systems of bases other than two. It is the purpose of this paper to carry this extension to its logical conclusion. The author describes how reflected systems of different bases may be composed. The equations for translating between the conventional and reflected systems are then derived. It is also demonstrated how the reflected binary system is a special case of reflected number systems and how the general case simplifies for the reflected binary case.*

[GH55] W. H. Gottschalk et G. A. Hedlund. *Topological Dynamics*. American Mathematical Society, Providence, Rhode Island. 1955.

[GH64] W. Gottschalk et G. Hedlund. *A Characterization of the Morse Minimal Set*. Proceedings of the American Mathematical Society, tome 15(1), pages 70–74, 1964. doi:10.1090/S0002-9939-1964-0158386-X

[GJ79] M. R. Garey et D. S. Johnson. *Computers and Intractability - A guide to the Theory of NP-Completeness*. W.H. Freeman and Company. 1979.

[GK72] Y. Gurevich et I. Koryakov. *Remarks on Berger's paper on the domino problem*. Siberian Math. Journal.

[GR97] D. Giammarresi et A. Restivo. *Two-dimensional languages*. Dans *Handbook of Formal Languages*, tome 3. Beyond Words. Springer. 1997.

[GRST96] D. Giammarresi, A. Restivo, S. Seibert et W. Thomas. *Monadic Second-Order Logic over Rectangular Pictures and Recognizability by Tiling Systems*. Information and Computation, tome 125, pages 32–45, 1996.

Résumé : *It is shown that a set of pictures (rectangular arrays of symbols) is recognized by a finite tiling system iff it is definable in existential monadic second-order logic. As a consequence, finite tiling systems constitute a notion of recognizability over two-dimensional inputs which at the same time generalizes finite-state recognizability over strings and also matches a natural logic. The proof is based on the Ehrenfeucht-Fraissé technique for first-order logic and an implementa- tion of "threshold counting" within tiling systems.*

[Grä89] E. Grädel. *Dominoes and the complexity of subclasses of logical theories*. Annals of Pure and Applied Logic, tome 43(1), pages 1–30, 1989. doi:10.1016/0168-0072(89)90023-7

Résumé : *The complexity of subclasses of logical theories (mainly Presburger and Skolem arithmetic) is studied. The subclasses are defined by the structure of the quantifier prefix.*
For this purpose finite versions of dominoes (tiling problems) are used. Dominoes were introduced in the sixties as a tool to prove the undecidability of the $\forall\exists\forall$*-case of the predicate calculus and have found in the meantime many other applications. Here it is shown that problems in complexity classes NTIME(T(n)) are reducible to domino problems where the space to be tiled is a square of size T(n). Because of their simple combinatorial structure these dominoes provide a convinient method for providing lower complexity bounds for simple formula classes in logical theories.*
Using this method it is shown that the class of $\exists\forall^\star$*-formulas in Presburger arithmetic has exponential complexity. This seems to be the simplest class with this property because the set of* $\exists^\star$*-sentences in Presburger arithmetic is NP-complete*

and the classes which is shown to be fixed prefixes (i.e. where also the number of variables is limited) are all contained in appropriate levels of the polynomial time-hierarchy.
Skolem arithmetic is the theory of positive natural numbers with multiplication and 's thus (isomorphic to) the weak direct power of Presburger arithmetic. For the theory in general as well as for most subclasses the complexity is one exponential step higher than in the case of Presburger arithmetic. An exception is the class of $\exists^\star$-formulas which is shown to be NP-complete. On the other hand there is a formula class with fixed dimension which already has an exponential lower complexity bound.
The last section mentions some results on other logical theories and indicates some possible lines of future research.

[GW04] M. Grohe et S. Wöhrle. *An existential locality theorem*. Annals of Pure and Applied Logic, tome 129(1-3), pages 131–148, 2004. doi:10.1016/j.apal.2004.01.005

Résumé : *We prove an existential version of Gaifmanś locality theorem and show how it can be applied algorithmically to evaluate existential first-order sentences in finite structures.*

[Han65] W. Hanf. *Model-theoretic methods in the study of elementary logic*. Dans *Symposium in the Theory of Models*, pages 132–145. 1965.

[Han74] W. Hanf. *Non Recursive Tilings of the Plane I*. Journal of Symbolic Logic, tome 39(2), pages 283–285, 1974. euclid.jsl/1183739044

[Har85] D. Harel. *Recurring Dominoes: Making the Highly Undecidable Highly Understandable*. Annals of Discrete Mathematics, tome 24, pages 51–72, 1985.

Résumé : *In recent years many diverse logical systems for reasoning about programs have been shown to possess a highly undecidable, viz Π^1_1-complete validity problem. All such known results are reproved in this paper in a uniform and transparent manner by reductions from* recurring domino problems*. These are simple variants of the classical unbounded domino (or tiling) problems introduced by Wang and the bounded versions defined by Lewis. While the former are (weakly) undecidable and the latter complete in various complexity classes, the problem in the new class are Σ^1_1-complete.*
It is hoped that the paper, which contains also NP-, PSPACE-, Π^0_1 and Π^0_2-hardness results for various logical systems, will enhance interest in the appealing medium of domino problems as a useful set of reduction tools for exhibiting 'bad behaviour'.

[Har86] D. Harel. *Effective transformations on infinite trees, with applications to high undecidability, dominoes, and fairness*. Journal of the ACM, tome 33(1), pages 224–248, 1986. doi:10.1145/4904.4993

Résumé : *Elementary translations between various kinds of recursive trees are presented. It is shown that trees of either finite or countably infinite branching can be effectively put into one-one correspondence with infinitely branching trees in such a way that the infinite paths of the latter correspond to the ϕ-abiding infinite paths of the former. Here ϕ can be any member of a very wide class of properties of infinite paths. For many properties ϕ, the converse holds too. Two of the applications involve (a) the formulation of large classes of highly undecidable variants of classical computational problems, and in particular, easily describable domino problems that are Π^1_1-complete, and (b) the existence of a general method for proving termination of nondeterministic or concurrent programs under any reasonable notion of fairness.*

[Hed69] G. A. Hedlund. *Endomorphisms and automorphisms of the shift dynamical system*. Theory of Computing Systems, tome 3(4), pages 320–375, 1969.

[Her71] H. Hermes. *A simplified proof for the unsolvability of the decision problem in the case $\wedge \vee \wedge$*. Dans *Logic Colloquium 1969*, *Studies in Logic and the Foundations of Mathematics*, tome 61, pages 307–310. 1971. doi:doi:10.1016/S0049-237X(08)71235-7

[HM10] M. Hochman et T. Meyerovitch. *A characterization of the entropies of multidimensional shifts of finite type*. Annals of Mathematics, tome 171(3), pages 2011–2038, 2010.

Résumé : *We show that the values of entropies of multidimensional shifts of finite type (SFTs) are characterized by a certain computation-theoretic property: a real number $h \geq 0$ is the entropy of such an SFT if and only if it is right recursively enumerable, i.e. there is a computable sequence of rational numbers converging to h from above. The same characterization holds for the entropies of sofic shifts. On the other hand, the entropy of a strongly irreducible SFTs is computable.*

[Hoc09a] M. Hochman. *A note on universality in multidimensional symbolic dynamics*. Discrete and Continuous Dynamical Systems S, tome 2(2). doi:10.3934/dcdss.2009.2.301

Résumé : *We show that in the category of effective $\mathbb{Z}$-dynamical systems there is a universal system, i.e. one that factors onto every other effective system. In particular, for $d \geq 3$ there exist d-dimensional shifts of ?nite type which are universal for 1-dimensional subactions of SFTs. On the other hand, we show that there is no universal effective $\mathbb{Z}^d$-system for $d \geq 2$, and in particular SFTs cannot be universal for subactions of rank ≥ 2. As a consequence, a decrease in entropy and Medvedev degree and periodic data are not sufficient for a factor map to exists between SFTs. We also discuss dynamics of cellular automata on their limit sets and show that (except for the unavoidable presence of a periodic point) they can model a large class of physical systems.*

[Hoc09b] M. Hochman. *On the dynamics and recursive properties of multidimensional symbolic systems*. Inventiones Mathematicae, tome 176(1), page 2009, 2009.

Résumé : *We study the (sub)dynamics of multidimensional shifts of finite type and sofic shifts, and the action of cellular automata on their limit sets. Such a subaction is always an effective dynamical system: i.e. it is isomorphic to a subshift over the Cantor set the complement of which can be written as the union of a recursive sequence of basic sets. Our main result is that, to varying degrees, this recursive-theoretic condition is also sufficient. We show that the class of expansive subactions of multidimensional sofic shifts is the same as the class of expansive effective systems, and that a general effective system can be realized, modulo a small extension, as the subaction of a shift of finite type or as the action of a cellular automaton on its limit set (after removing a dynamically trivial set). As applications, we characterize, in terms of their computational properties, the numbers which can occur as the entropy of cellular automata, and construct SFTs and CAs with various interesting properties.*

[Hoo66] P. K. Hooper. *The Undecidability of the Turing Machine Immortality Problem*. Journal of Symbolic Logic, tome 31(2), pages 219–234, 1966.

[Hou08] A. O. Houcine. *On finitely generated models of theories with at most countably many nonisomorphic finitely generated models*. Preprint. 2008.

Résumé : *We study finitely generated models of countable theories, having at most countably many nonisomorphic finitely generated models. We introduce a notion of rank of finitely generated models and we prove, when T has at most countably many nonisomorphic finitely generated models, that every finitely generated model has an ordinal rank. This rank is used to give a property of finitely* *generated models analogue to the Hopf property of groups and also to give a necessary and sufficient condition for a finitely generated model to be prime of its complete theory. We investigate some properties of limit groups of equationally noetherian groups, in respect to their ranks.*

[II96] K. C. II. *An aperiodic set of 13 Wang tiles*. Discrete Mathematics, tome 160, pages 245–251, 1996.

Résumé : *A new aperiodic tile set containing only 13 tiles over 5 colors is presented. Its construction is based on a recent technique developed by Kari. The tilings simulate the behavior of sequential machines that multiply real numbers in balanced representations by real constants.*

[Imm88] N. Immerman. *Nondeterministic space is closed under complementation*. SIAM Journal on Computing, tome 17(5), pages 935–938, 1988.

[JM97] A. Johnson et K. Madden. *Putting the Pieces Together: Understanding Robinson's Nonperiodic Tilings*. The College Mathematics Journal, tome 28(3), pages 172–181, 1997.

[JS72a] C. G. Jockusch et R. I. Soare. *Degrees of members of Π^0_1 classes*. Pacific J. Math., tome 40(3), pages 605–616, 1972. euclid.pjm/1102968559

[JS72b] N. D. Jones et A. L. Selman. *Turing machines and the spectra of first-order formulas with equality*. Dans *STOC '72: Proceedings of the fourth annual ACM symposium on Theory of computing*, pages 157–167. ACM, New York, NY, USA, 1972. doi:10.1145/800152.804909

[JS72c] C. G. J. Jr et R. I. Soare. $\prod^0_1$ *classes and degrees of theories*. Transactions of the American Mathematical Society, tome 173, pages 33–56, 1972.

Résumé : *Using the methods of recursive function theory we derive several results about the degrees of unsolvability of members of certain $\prod^0_1$ classes of functions (i.e. degrees of branches of certain recursive trees). As a special case we obtain information on the degrees of consistent extensions of axiomatizable theories, in particular effectively inseparable theories such as Peano arithmetic,* $\mathbf{P}$. *For example: THEOREM 1.* If a degree $\mathbf{a}$ contains a complete extension of $\mathbf{P}$, then every countable partially ordered set can be embedded in the ordering of degrees $\leq \mathbf{a}$. *(This strengthens a result of Scott and Tennenbaum that no such degree* $\mathbf{a}$ *is a minimal degree.) THEOREM 2.* If $\mathbf{T}$ is an axiomatizable, essentially undecidable theory, and if $\{a_n\}$ is a countable sequence of nonzero degrees, then $\mathbf{T}$ has continuum many complete extensions whose degrees are pairwise incomparable and incomparable with each a_n. *THEOREM 3.* There is a complete extension $\mathbf{T}$ of $\mathbf{P}$ such that no nonrecursive arithmetical set is definable in $\mathbf{T}$. *THEOREM 4.* There is an axiomatizable, essentially undecidable theory $\mathbf{T}$ such that any two distinct complete extensions of $\mathbf{T}$ are Turing incomparable. *THEOREM 5.* The set of degrees of consistent extensions of $\mathbf{P}$ is meager and has measure zero.

[Kar92] J. Kari. *The Nilpotency Problem of One-Dimensional Cellular Automata*. SIAM Journal on Computing, tome 21(3), pages 571–586, 1992.

Résumé : *The limit set of a celullar automaton consists of all the configurations of the automaton that can appear after arbitrarily long computations. It is known that the limit set is never empty – it contains at least one homogeneous configuration. A CA is called nilpotent if its limit set contains just one configuration. The present work proves that it is algorithmically undecidable whether a given one-dimensional cellular automaton is nilpotent. The proof is based on a generalization of the well-known result about the undecidability of the tiling problem of* *the plane. The generalization states that the tiling problem remains undecidable even if one considers only so-called NW-deterministic tile sets, that is, tile sets in which the left and upper neighbors of each tile determine the tile uniquely. The nilpotency problem is known to be undecidable for d-dimensional CA for $d \geq 2$. The result is the basis of the proof of Rice's theorem for CA limit sets, which states that every nontrivial property of limit sets is undecidable.*

[Kar94] J. Kari. *Reversibility and surjectivity problems of cellular automata*. Journal of Computer and System Sciences, tome 48(1), pages 149–182, 1994. doi:10.1016/S0022-0000(05)80025-X

Résumé : *The problem of deciding if a given cellular automaton (CA) is reversible (or, equivalently, if its global transition function is injective) is called the reversibility problem of CA. In this article we show that the reversibility problem is undecidable in case of two-dimensional CA. We also prove that the corresponding surjectivity problem – the problem of deciding if the global function is surjective – is undecidable for two-dimensional CA. Both problems are known to be decidable in case of one-dimensional CA. The proofs of the theorems are based on reductions from the well-known tiling problem of the plane, known also as the domino problem.*

[Kar96] J. Kari. *A small aperiodic set of Wang tiles*. Discrete Mathematics, tome 160, pages 259–264, 1996.

Résumé : *A new aperiodic tile set containing only 14 Wang tiles is presented. The construction is based on Mealy machines that multiply Beatty sequences of real numbers by rational constants*

[Kar07] J. Kari. *The Tiling Problem Revisited* . Dans *Machines, Computations, and Universality (MCU)*, numéro 4664 dans Lecture Notes in Computer Science, pages 72–79. 2007. doi:10.1007/978-3-540-74593-8_6

Résumé : *We give a new proof for the undecidability of the tiling problem. Then we show how the proof can be modified to demonstrate the undecidability of the tiling problem on the hyperbolic plane, thus answering an open problem posed by R.M. Robinson in 1971*

[Kec95] A. S. Kechris. *Classical descriptive set theory*, *Graduate Texts in Mathematics*, tome 156. Springer-Verlag, New York. 1995. ISBN 0-387-94374-9

[KL05] D. Kuske et M. Lohrey. *Logical aspects of Cayley-graphs: the group case*. Annals of Pure and Applied Logic, tome 131(1-3), pages 263–286, 2005.

Résumé : *We prove that a finitely generated group is context-free whenever its Cayley-graph has a decidable monadic second-order theory. Hence, by the seminal work of Muller and Schupp, our result gives a logical characterization of context-free groups and also proves a conjecture of Schupp. To derive this result, we investigate general graphs and show that a graph of bounded degree with a high degree of symmetry is context-free whenever its monadic second-order theory is decidable. Further, it is shown that the word problem of a finitely generated group is decidable if and only if the first-order theory of its Cayley-graph is decidable.*

[KMW62] A. Kahr, E. F. Moore et H. Wang. *Entscheidungsproblem reduced to the* $\forall\exists\forall$ *case*. Proceedings of the National Academy of Sciences of the United States of America, tome 48(3), pages 365–377, 1962.

[Knu05] D. E. Knuth. *Generating all Tuples and Permutations*, *The Art of Computer Programming*, tome 4 fasc. 2. Addison-Wesley. 2005.

[KO08] J. Kari et N. Ollinger. *Periodicity and Immortality in Reversible Computing*. Dans *MFCS 2008*, pages 419–430. 2008.

Résumé : *We investigate the decidability of the periodicity and the immortality problems in three models of reversible computation: reversible counter machines, reversible Turing machines and reversible one-dimensional cellular automata. Immortality and periodicity are properties that describe the behavior of the model starting from arbitrary initial configurations: immortality is the property of having at least one non-halting orbit, while periodicity is the property of always eventually returning back to the starting configuration. It turns out that periodicity and immortality problems are both undecidable in all three models. We also show that it is undecidable whether a (not-necessarily reversible) Turing machine with moving tape has a periodic orbit.*

[KP99] J. Kari et P. Papasoglu. *Deterministic Aperiodic Tile Sets*. Geometric And Functional Analysis, tome 9, pages 353–369, 1999.

Résumé : *Wang tiles are square tiles with colored edges. We construct an aperiodic set of Wang tiles that is strongly deterministic in the sense that any two adjacent edges of a tile determine the tile uniquely. Consequently, the tiling group of this set is not hyperbolic and it acts discretely and co-compactly on a CAT(0) space.*

[Kre53] G. Kreisel. *A Variant to Hilbert's Theory of the Foundations of Arithmetic*. The British Journal for the Philosophy of Science, tome 4(14), pages 107–129, 1953.

Résumé : *In Hilbert's theory of the foundations of any given branch of mathematics the main problem is to establish the consistency (of a suitable formalisation) of this branch. Since the (intuitionist) criticisms of classical logic, which Hilbert's theory was intended to meet, never even alluded to inconsistencies (in classical arithmetic), and since the investigations of Hilbert's school have always established much more than mere consistency, it is natural to formulate another general problem in the foundations of mathematics: to translate statements of theorems and proofs in the branch considered into those of some preferred system, where the translation must satisfy certain appropriate conditions (interpretation). The problem is relative to the choice of preferred system, as is Hilbert's consistency problem since he required the consistency to be established by particular methods (finitist ones). A finitist interpretation of classical number theory, which has been published in full detail elsewhere, is here described by means of typical examples. Partial results on analysis (theory of arbitrary functions whose arguments and values are the non-negative integers) are here presented for the first time. One of these results is restricted to functions whose values are bounded ; its interest derives from the fact that real numbers may be represented by such functions. It is hoped that diverse general observations and comments, which would bore the specialist, may be of help to the general reader. The specialist may find some points of interest in the last two sections of the main text and in the notes following it.*

[Kur66] K. Kuratowski. *Topology, Vol. I, 3rd edition*. NY: Academic Press. 1966.

[Lev73] L. A. Levin. *Universal search problems*. Problemy Peredachi Informatsii, tome 9(3), pages 115–116, 1973. traduction dans [Tra84].

Résumé : *Several well-known problems of the "search" type are discussed, and it is proved that those problems can be solved only in the time it takes to solve any problems of the indicated type in general.*

[Lev84] L. A. Levin. *Problems, Complete in "Average" Instance*. Dans *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing (STOC)*, page 465. ACM, 1984. doi:10.1145/800057.808713

[Lev86] L. A. Levin. *Average Case Complete Problems*. SIAM Journal on Computing, tome 15(1), pages 285–286, 1986. doi:10.1137/0215020

Résumé : *Many interesting combinatorial problems were found to be NP-complete. Since there is little hope to solve them fast in the worst case, researchers look for algorithms which are fast just "on average". This matter is sensitive to the choice of a particular NP-complete problem and a probability distribution of its instances. Some of these tasks were easy and some not. But one needs a way to distinguish the "difficult on average" problems. Such negative results could not only save "positive" efforts but may also be used in areas (like cryptography) where hardness of some problems is a frequent assumption. It is shown below that the Tiling problem with uniform distribution of instances has no polynomial "on average" algorithm, unless every NP-problem with every simple probability distribution has it. It is interesting to try to prove similar statements for other NP-problems which resisted so far "average case" attacks.*

[Lew78] H. R. Lewis. *Complexity of solvable cases of the decision problem for the predicate calculus*. Dans *Symposium on Foundations of Computer Science (FOCS)*, pages 35–47. IEEE Computer Society, Los Alamitos, CA, USA, 1978. doi:10.1109/SFCS.1978.9

Résumé : *We analyze the computational complexity of determining whether F is satisfiable when F is a formula of the classical predicate calculus which obeys certain syntactic restrictions. For example, for the monadic predicate calculus and the Gödel or $\exists \ldots \exists\forall\forall\exists \ldots \exists$ prefix class we obtain lower and upper nondeterministic time bounds of the form $cn/\log n$. The main tool in in these proofs is a finite version of Wang's domino problem, about which we present an interesting open question.*

[Lew79] H. R. Lewis. *Unsolvable Classes of Quantificational Formulas*. Addison-Wesley. 1979. ISBN 0-201-04069-7

[Lib04] L. Libkin. *Elements of Finite Model Theory*. Texts in Theoretical Computer Science, an EATCS Series. Springer Verlag. 2004. ISBN 3540212027

[Lin96] D. A. Lind. *A zeta function for $\mathbb{Z}^d$-actions*. Dans *Ergodic Theory of $\mathbb{Z}^d$-actions* (M. Pollicott et K. Schmidt, réds.), *LMS Lecture Note Series*, tome 228, pages 433–450. Cambridge University Press, 1996. ISBN 0521576881

Résumé : *We define a zeta function for $\mathbb{Z}^d$ actions α that generalizes the Artin-Mazur zeta function for a single transformation. This zeta function is a conjugacy invariant, can be computed explicitely in some cases, and has a product formula over finite orbits. The analytic behavior of the zeta function for $d \geq 2$ is quite different from the case $d = 1$. Even for higher-dimensional actions of finite type the zeta function is typically transcendental and has natural boundary a circle of finite radius. We compute the radius of convergence of the zeta function for a class of algebraic $\mathbb{Z}^d$-actions. We conclude by conjecturing a general description of the analytic behavior of these zeta functions and discussing some further problems.*

[Lin04] D. A. Lind. *Multi-Dimensional Symbolic Dynamics*. Dans *Symbolic Dynamics and its Applications* (S. G. Williams, réd.), numéro 60 dans Proceedings of Symposia in Applied Mathematics, pages 61–79. American Mathematical Society, 2004.

[LM95] D. A. Lind et B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, New York, NY, USA. 1995. ISBN 0521551242

[LP98] H. R. Lewis et C. H. Papadimitriou. *Elements of the Theory of Computation*. Prentice-Hall. 1998. ISBN 0132624788

[LW07] G. Lafitte et M. Weiss. *Universal tilings*. Dans *STACS*, pages 367–380. 2007. doi:10.1007/978-3-540-70918-3_32

Résumé : *Wang tiles are unit size squares with colored edges. To know if a given finite set of Wang tiles can tile the plane while respecting colors on edges is undecidable. Bergerś proof of this result shows the equivalence between tilings and Turing machines and thus tilings can be seen as a computing model. We thus have tilings that are Turing-universal, but there lacks a proper notion of universality for tilings. In this paper, we introduce natural notions of universality and completeness for tilings. We construct some universal tilings and try to make a first hierarchy of tile sets with a universality criteria.*

[LW08] G. Lafitte et M. Weiss. *Computability of Tilings*. Dans *IFIP*, *IFIP International Federation for Information Processing*, tome 273, pages 187–201. Springer Boston, 2008.

Résumé : *Wang tiles are unit size squares with colored edges. To know whether a given finite set of Wang tiles can tile the plane while respecting colors on edges is undecidable. Robinson's tiling is an auto-similar tiling in which the computation of a Turing machine can be carried out. By using this construction and by considering a strong notion of simulation between tilings, we prove computability results for tilings. In particular, we prove theorems on tilings that are similar to Kleene's recursion theorems. Then we define and show how to construct reductions between sets of tile sets. We generalize this construction to be able to transform a tile set with a given recursively enumerable property into a tile set with another property. These reductions lead naturally to a Rice-like theorem for tilings.*

[Mak74] J. A. Makowsky. *On some conjectures connected with complete sentences*. Fundamenta Mathematicae, tome 81, pages 193–202, 1974.

Résumé : *Some conjectures on the finite axiomatizability of complete* $\aleph_1$*-categorical theories are discussed and related to open problems in group theory. A problem of Chang and Keisler is solved.*

[Mat98a] O. Matz. *On piecewise testable, starfree, and recognizable picture languages*. Dans *Foundations of Software Science and Computation Structures (FoSSaCs)*, *Lecture Notes in Computer Science*, tome 1378, pages 203–210. Springer, 1998. doi:10.1007/BFb0053551

Résumé : *We isolate a technique for showing that a picture language (i.e. a two-dimensional language) is not recognizable. Then we prove the non-recognizability of a picture language that is both starfree (i.e., definable by means of union, concatenation, and complement) and piecewise testable (i.e., definable by means of allowed subpictures), solving an open question in [GR96]. We also define local, locally testable, and locally threshold testable picture languages and summarize known inclusion results for these classes. The classes of piecewise testable, locally testable, and locally threshold testable picture languages can, as in the word case, be characterized by certain (fragments of) first-order logics.*

[Mat98b] O. Matz. *One Quantifier Will Do in Existential Monadic Second-Order Logic over Pictures*. Dans *Proceedings of the 23rd International Symposium on Mathematical Foundations of Computer Science, MFCS*, pages 751–759. 1998.

Résumé : *We show that every formula of the existential fragment of monadic second-order logic over picture models (i.e., finite, two-dimensional, coloured grids) is equivalent to one with only one existential monadic quantifier. The corresponding claim is true for the class of word models ([Tho82]) but not for the* *class of graphs ([Ott95]). The class of picture models is of particular interest because it has been used to show the strictness of the different (and more popular) hierarchy of quantifier alternation.*

[Mey05] A. Meyer. *Graphes infinis de présentation finie*. Thèse de doctorat, Université de Rennes 1. 2005.

[Mey10] T. Meyerovitch. *Growth-type invariants for $\mathbb{Z}^d$ subshifts of finite type and arithmetical classes of real numbers*. Inventiones Mathematicae.

Résumé : *We discuss some numerical invariants of multidimensional shifts of finite type (SFTs) which are associated with the growth rates of the number of admissible finite configurations. Extending an unpublished example of Tsirelson, we show that growth complexities of the form $\exp(n^\alpha)$ are possible for non-integer α's. In terminology of Carvalho, such subshifts have entropy dimension α. The class of possible α's are identified in terms of arithmetical classes of real numbers of Weihrauch and Zheng.*

[MH38] M. Morse et G. A. Hedlund. *Symbolic Dynamics*. American Journal of Mathematics, tome 60(4), pages 815–866, 1938.

[Mil11] J. S. Miller. *Two Notes on subshifts*. Proceedings of the American Mathematical Society.

Résumé : *We prove two unrelated results about subshifts. First, we give a condition on the lengths of forbidden words that is sufficient to guarantee that the corresponding subshift is nonempty. The condition implies that, for example, any sequence of binary words of lengths 5, 6, 7, ... is avoidable. As another application, we derive a result of Durand, Levin and Shen [2, 3] that there are infinite sequences such that every substring has high Kolmogorov complexity. In particular, for any $d < 1$, there is a $b \in \mathbb{N}$ and an infinite binary sequence X such that if τ is a substring of X, then τ has Kolmogorov complexity greater than $d|\tau| - b$. The second result says that, from the standpoint of computability theory, any behavior possible from an arbitrary effectively closed subset of $n^{\mathbb{N}}$ (i.e., a Π^0_1 class) is exhibited by an effectively closed subshift. In technical terms, every Π^0_1 Medvedev degree contains a Π^0_1 subshift. This answers a question of Simpson [10].*

[Mor21] H. M. Morse. *Recurrent Geodesics on a Surface of Negative Curvature*. Transactions of the American Mathematical Society, tome 22(1), pages 84–100, 1921. doi:10.1090/S0002-9947-1921-1501161-8

[Moz89] S. Mozes. *Tilings, substitutions systems and dynamical systems generated by them*. J. d'Analyse Math., tome 53, pages 139–186, 1989.

Résumé : *The object of this work is to study the properties of dynamical systems defined by tilings. A connection to symbolic dynamical systems defined by one- and two-dimensional substitution systems is shown. This is used in particuliar to show the existence of a tiling system such that its corresponding dynamical system is minimal and topological weakly mixing. We remark that for one-dimensional tilings the dynamical system always contains periodic points.*

[MSU03] A. Muchnik, A. Semenov et M. Ushakov. *Almost periodic sequences*. Theoretical Computer Science, tome 304(1-3), pages 1–33, 2003. doi:10.1016/S0304-3975(02)00847-2

Résumé : *This paper studies properties of almost periodic sequences (also known as uniformly recursive). A sequence is almost periodic if for every finite string that occurs infinitely many times in the sequence there exists a number m such that every segment of length m contains an occurrence of the word. We study closure properties of the set of almost periodic sequences, ways to generate such sequences (including a general way), computability issues and Kolmogorov complexity of prefixes of almost periodic sequences.*

[Mye74] D. Myers. *Non Recursive Tilings of the Plane II*. Journal of Symbolic Logic, tome 39(2), pages 286–294, 1974. euclid.jsl/1183739045

[Oge] F. Oger. *Tilings and associated relational structures*. Http://arxiv.org/abs/0902.3429.

[Oge04] F. Oger. *Algebraic and model-theoretic properties of tilings*. Theoretical Computer Science, tome 319, pages 103–126, 2004.

Résumé : *We investigate the relations between the geometric properties of tilings and the algebraic and model-theoretic properties of associated relational structures. Isomorphism and local isomorphism of tilings up to translation correspond to isomorphism and elementary equivalence of relational structures. In particular, two Penrose tilings, or two Robinson tilings, are elementarily equivalent. Classical results concerning the local isomorphism property and the extraction preorder for tilings are generalized to uniformly locally finite relational structures.*
Then, we define equational structures, which generalize both Cayley graphs of groups and relational structures associated to tilings, and for which we have an appropriate notion of free structure relative to a system of equations. For each finite system Σ of prototiles and local configurations, we give a finite set of local conditions characterizing the connected relational structures which are homomorphic images of Σ-tilings. It follows that tilings are free relative to finite systems of equations which express these conditions. We also prove that the theory of a tiling is superstable, model-complete, and can be axiomatized by $\forall\exists$ sentences. One $\forall\exists$ sentence suffices in the case of Penrose tilings or Robinson tilings.

[Oll08] N. Ollinger. *Two-by-Two Substitution Systems and the Undecidability of the Domino Problem*. Dans *CiE 2008*, numéro 5028 dans Lecture Notes in Computer Science, pages 476–485. 2008.

Résumé : *Thanks to a careful study of elementary properties of two-by- two substitution systems, we give a complete self-contained elementary construction of an aperiodic tile set and sketch how to use this tile set to elementary prove the undecidability of the classical Domino Problem.*

[Pav11] R. Pavlov. *A class of nonsofic multidimensional shift spaces*. Proceedings of the American Mathematical Society.

[Poi80] B. Poizat. *Une théorie finiement axiomatisable et superstable*. Groupe d'études de théories stables, tome 3, pages 1–9, 1980.

[Rob71] R. M. Robinson. *Undecidability and Nonperiodicity for Tilings of the Plane*. Inventiones Mathematicae, tome 12(3). doi:10.1007/BF01418780

[RU06] A. Rumyantsev et M. Ushakov. *Forbidden Substrings, Kolmogorov Complexity and Almost Periodic Sequences*. Dans *23rd International Symposium on Theoretical Aspects of Computer Science, STACS 2006*, numéro 3884 dans LNCS, pages 396–407. Springer-Verlag, 2006. doi:10.1007/11672142_32

Résumé : *Assume that for some $\alpha < 1$ and for all nutural n a set F_n of at most $2^{\alpha n}$ forbidden binary strings of length n is fixed. Then there exists an infinite binary sequence ω that does not have (long) forbidden substrings. We prove this combinatorial statement by translating it into a state- ment about Kolmogorov complexity and compare this proof with a com- binatorial one based on Laslo Lovasz local lemma. Then we construct an almost periodic sequence with the same property (thus combines the results from [1] and [2]). Both the combinatorial proof and Kolmogorov complexity argument can be generalized to the multidimensional case.*

[RW92] C. Radin et M. Wolff. *Space Tilings and Local Isomorphism*. Geometriae Dedicata, tome 42, pages 355–360, 1992.

Résumé : *We prove for a large class of tilings that, given a finite tile set, if it is possible to tile Euclidean n-space with isometric copies of this set, then there is a tiling with the 'local isomorphism property'.*

[Sad08] L. Sadun. *Topology of Tiling Spaces*. Numéro 46 dans University Lecture Series. American Mathematical Society. 2008.

[Sal89] O. Salon. *Quelles tuiles ! (Pavages apériodiques du plan et automates bidimensionnels)*. Journal de Théorie des Nombres de Bordeaux, tome 1(1), pages 1–26, 1989.

Résumé : *La récente découverte des quasicristaux et leurs liens avec les pavages de Penrose ont entraîné un regain d'intérêt pour les pavages apériodiques du plan. Nous montrons ici que le pavage régulier de Robinson est engendré par un automate fini bidimensionnel, et qu'il donne une généralisation à deux dimensions du pliage de papier.*

[Sal10] P. V. Salimov. *On Uniform Recurrence of a Direct Product*. Discrete Mathematics and Theoretical Computer Science, tome 12(4), pages 1–8, 2010.

Résumé : *The direct product of two words is a naturally defined word on the alphabet of pairs of symbols. An infinite word is uniformly recurrent if each its subword occurs in it with bounded gaps. An infinite word is strongly recurrent if the direct product of it with each uniformly recurrent word is also uniformly recurrent. We prove that fixed points of the expanding binary symmetric morphisms are strongly recurrent. In particular, such is the Thue-Morse word.*

[SB99] T. Schwentick et K. Barthelmann. *Local Normal Forms for First-Order Logic with Applications to Games and Automata*. Discrete Mathematics and Theoretical Computer Science, tome 3, pages 109–124, 1999.

Résumé : *Building on work of Gaifman [Gai82] it is shown that every first-order formula is logically equivalent to a formula of the form $\exists x1 \ldots x_l \forall y \phi$ where ϕ is r-local around y, i. e. quantification in ϕ is restricted to elements of the universe of distance at most r from y.*
From this and related normal forms, variants of the Ehrenfeucht game for first-order and existential monadic second-order logic are developed that restrict the possible strategies for the spoiler, one of the two players. This makes proofs of the existence of a winning strategy for the duplicator, the other player, easier and can thus simplify inexpressibility proofs.
As another application, automata models are defined that have, on arbitrary classes of relational structures, exactly the expressive power of first-order logic and existential monadic second-order logic, respectively.

[Sch00] K. Schmidt. *Multi-Dimensional Symbolic Dynamical Systems*. Rapport technique ESI 870, Erwin Schrodinger International Institute for Mathematical Physics. 2000.

Résumé : *The purpose of this note is to point out some of the phenomena which arise in the transition from classical shifts of finite type $X \subset A^{\mathbb{Z}}$ to* multi-dimensional *shifts of finite type $X \subset A^{\mathbb{Z}^d}, d \geq 2$, where A is a finite alphabet. We discuss rigidity properties of certain multi-dimensional shifts, such as the appearance of an unexpected intrinsic algebraic structure or the scarcity of isomorphisms and invariant measures. The final section concentrates on group shifts with finite or uncountable alphabets, and with the symbolic representation of such shifts in the latter case.*

[See91] D. Seese. *The structure of the models of decidable monadic theories of graphs*. Annals of pure and applied logic, tome 53(2), pages 169–195, 1991.

> **Résumé :** *In this article the structure of the models of decidable (weak) monadic theories of planar graphs is investigated. It is shown that if the (weak) monadic theory of a class K of planar graphs is decidable, then the tree-width in the sense of Robertson and Seymour (1984) of the elements of K is universally bounded and there is a class T of trees such that the (weak) monadic theory of K is interpretable in the (weak) monadic theory of T.*

[Sho60] J. Shoenfield. *Degrees of Models*. Journal of Symbolic Logic, tome 25(3), pages 233–237, 1960.

[Sim11a] S. G. Simpson. *Mass Problems Associated with Effectively Closed Sets*. In preparation. 2011.

[Sim11b] S. G. Simpson. *Medvedev Degrees of 2-Dimensional Subshifts of Finite Type*. Ergodic Theory and Dynamical Systems.

> **Résumé :** *In this paper we apply some fundamental concepts and results from recursion theory in order to obtain an apparently new counterexample in symbolic dynamics. Two sets X and Y are said to be Medvedev equivalent if there exist partial recursive functionals from X into Y and vice versa. The Medvedev degree of X is the equivalence class of X under Medvedev equivalence. There is an extensive recursion-theoretic literature on the lattice of Medvedev degrees of nonempty* Π^1_0 *subsets of* $\{0,1\}^{\mathbb{N}}$. *This lattice is known as* P_s . *We prove that* P_s *consists precisely of the Medvedev degrees of 2-dimensional subshifts of finite type. We use this result to obtain an infinite collection of 2-dimensional subshifts of finite type which are, in a certain sense, mutually incompatible.*

[Tho91] W. Thomas. *On logics, tilings, and automata*. Dans *Proceedings of the 18th ICALP* (S. Berlin, réd.), *LNCS*, tome 510, pages 441–454. 1991. doi:10.1007/3-540-54233-7_154

> **Résumé :** *We relate the logical and the automata theoretic approach to define sets of words, trees, and graphs. For this purpose a notion of graph acceptor is introduced which can specify monadic second-order properties and allows to treat known types of finite automata in a common framework. In the final part of the paper, we discuss infinite graphs that have a decidable monadic second-order theory.*

[Tho97] W. Thomas. *Languages, Automata, and Logic*. Dans *Handbook of Formal Languages* (G. Rosenberg et A. Salomaa, réds.), tome 3. Beyond Words. Springer. 1997.

[Thu12] A. Thue. *Über die gegenseitige Lage gleicher Teiler gewisser Zeichenreihen*. Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl., tome 12. traduction dans [Ber95].

[Tra84] B. A. Trakhtenbrot. *A survey of Russian Approached to* Perebor *(Brute-Force Search) Algorithms*. Annals of the History of Computing, tome 6(4), pages 384–400, 1984. doi:10.1109/MAHC.1984.10036

[vEB97] P. van Emde Boas. *The convenience of Tilings*. Dans *Complexity, Logic, and Recursion Theory*, *Lecture Notes in Pure and Applied Mathematics*, tome 187. CRC. 1997. ISBN 0824700260

> **Résumé :** *Tiling problems provide for a very simple and transparent mechanism for encoding machine computations. This gives rise to rather simple master reductions showing various versions of the tiling problem complete for various complexity classes. We investigate the potential for using these tiling problems in subsequent reductions showing hardness of the combinatorial problems that really matter.*

We ilustrate our approach by means of three examples: a short reduction chain to the Knapsack problem followed by a Hilbert 10 reduction using similar ingredients. Finally we reprove the Deterministic Exponential Time lowerbound for satis ablility in Propositional Dynamic Logic.
The resulting reductions are relatively simple ; they do however infringe on the principle of orthogonality of reductions since they abuse extra structure in the instances of the problems reduced from which results from the fact that these instances were generated by a master reduction previously.

[Wan60] H. Wang. *Proving Theorems by Pattern Recognition I*. Communications of the ACM, tome 3(4), pages 220–234, 1960.

[Wan61] H. Wang. *Proving theorems by Pattern Recognition II*. Bell Systems technical journal, tome 40, pages 1–41, 1961.

Résumé : *Theoretical questions concerning the possibilities of proving theorems by machines are considered here from the viewpoint that emphasizes the underlying logic. A proof procedure for the predicate calculus is given that contains a few minor peculiar features. A fairly extensive discussion of the decision problem is given, including a partial solution to the $\forall x \exists y \forall z$ satisfiability case, an alternative procedure ofr the $\forall x \forall y \exists z$ case and a rather detailed treatment of Skolem's case. In connection with the $\forall x \exists y \forall z$ case, an amusing combinatorial problem is suggested in Section 4.1. Some simple mathematical examples are considered in Section VI.*

[Wan63] H. Wang. *Dominoes and the $\forall\exists\forall$ case of the decision problem*. Mathematical Theory of Automata, pages 23–55.

[Wei73] B. Weiss. *Subshifts of finite type and sofic systems*. Monatshefte für Mathematik, tome 77, pages 462–474, 1973.

[Wei00] K. Weihrauch. *Computable analysis*. Springer. 2000. ISBN 978-3-540-66817-6

Annexe C

STRUCTURAL ASPECTS OF TILINGS

L'article [I7] a été écrit en collaboration avec Alexis Ballier et Bruno Durand. Il étudie la relation d'ordre $\prec$ définie au chapitre 2

Abstract

In this paper, we study the structure of the set of tilings produced by any given tile-set. For better understanding this structure, we address the set of finite patterns that each tiling contains.

This set of patterns can be analyzed in two different contexts: the first one is combinatorial and the other topological. These two approaches have independent merits and, once combined, provide somehow surprising results.

The particular case where the set of produced tilings is countable is deeply investigated while we prove that the uncountable case may have a completely different structure.

We introduce a pattern preorder and also make use of Cantor-Bendixson rank. Our first main result is that a tile-set that produces only periodic tilings produces only a finite number of them. Our second main result exhibits a tiling with exactly one vector of periodicity in the countable case.

C.1 Introduction

Tilings are basic models for geometric phenomena of computation: local constraints they formalize have been of broad interest in the community since they capture geometric aspects of computation [Rob71, Ber64, Han74, Mye74, DLS08]. This phenomenon was discovered in the sixties when tiling problems happened to be crucial in logic: more specifically, interest shown in tilings drastically increased when Berger proved the undecidability of the so-called domino problem [Ber64] (see also [GK72] and the well known book [BGG01] for logical aspects). Later, tilings were basic tools for complexity theory (see the nice review of Peter van Emde Boas [vEB97] and some of Leonid Levin's paper such as [Lev86]).

Because of growing interest for this very simple model, several research tracks were aimed directly on tilings: some people tried to generate the most complex tilings with the most simple constraints (see [Rob71, Han74, Mye74, DLS08]), while others were most interested in structural aspects (see [RW92, Dur99]).

In this paper we are interested in structural properties of tilings. We choose to focus on finite patterns tilings contain and thus introduce a natural preorder on tilings: a tiling is extracted from another one if all finite patterns that appear in the first one also appear in the latter. We develop this combinatorial notion in Section C.2.1. This approach can be expressed in terms of topology (subshifts of finite type) and we shall explain the relations between both these approaches in Section C.2.2.

It is important to stress that both these combinatorial and topological approaches have independent merits. Among the results we present, different approaches are indeed used for proofs.

More specifically, our first main result (Theorem C.8) states that if a tile-set produces only periodic tilings then it produces only finitely many of them; despite its apparent simplicity, we did not find any proof of Theorem C.8 in the literature. Our other main result (Theorem C.11) which states that in the countable case a tiling with exactly one vector of periodicity exists is proved with a strong help of topology.

Our paper is organized as follows: Section C.2 is devoted to definitions (combinatorics, topology) and basic structural remarks. In Section C.3 we prove the existence of minimal and maximal elements in tilings enforced by a tile-set. Then we present an analysis in terms of Cantor-Bendixson derivative which provides powerful tools. We study the particular case where tilings are countable and present our main results. We conclude by some open problems.

C.2 Definitions

C.2.1 Tilings

We present notations and definitions for tilings since several models are used in literature: Wang tiles, geometric frames of rational coordinates, local constraints... All these models are equivalent for our purposes since we consider very generic properties of them (see [CD04] for more details and proofs). We focus our study on tilings of the plane although our results hold in higher dimensions.

In our definition of tilings, we first associate a state to each cell of the plane. Then we impose a local constraint on them. More formally, Q is a finite set, called the *set of states*. A *configuration* c consists of cells of the plane with states, thus c is an element of $Q^{\mathbb{Z}^2}$. We denote by $c_{i,j}$ or $c(i,j)$ the state of c at the cell (i,j).

A tiling is a configuration which satisfies a given finite set of finite constraints everywhere. More specifically we express these constraints as a set of allowed patterns: a configuration is a tiling if around any of its cells we can see one of the allowed patterns:

Definition C.1 (*patterns*)

A *pattern* P is a finite restriction of a configuration *i.e.* an element of Q^V for some finite domain V of $\mathbb{Z}^2$. A pattern *appears* in a configuration c (resp. in some other pattern P') if it can be found somewhere in c (resp. in P'); *i.e.* if there exists a vector $t \in \mathbb{Z}^2$ such that $c(x+t) = P(x)$ on the domain of P (resp. if $P'(x+t)$ is defined for $x \in V$ and $P'(x+t) = P(x)$) .

By language extension we say that a pattern is *absent* or *omitted* in a configuration if it does not appear in it.

Definition C.2 (*tile-sets and tilings*)

A *tile-set* is a tuple $\tau = (Q, \mathcal{P}_\tau)$ where $\mathcal{P}_\tau$ is a finite set of patterns on Q. All the elements of $\mathcal{P}_\tau$ are supposed to be defined on the same domain denoted by V ($\mathcal{P}_\tau \subseteq Q^V$).

A *tiling* by τ is a configuration c equal to one of the patterns on all cells:

$$\forall x \in \mathbb{Z}^2, c|_{V+x} \in \mathcal{P}_\tau$$

We denote by $\mathcal{T}_\tau$ the set of tilings by τ.

Notice that in the definition of one tile-set we can allow patterns of different definition domains provided that there are a finite number of them.

An example of a tile-set defined by its allowed patterns is given in Fig. C.1. The produced tilings are given in Fig. C.2; the meaning of the edges in the graph will be explained later; tilings are represented modulo shift. In A_i and B_i, i is an integer that represents the size of the white stripe.

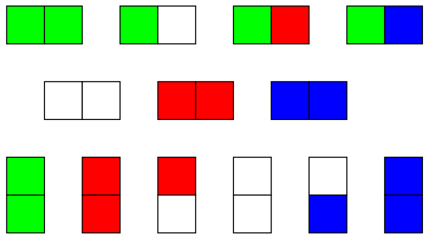
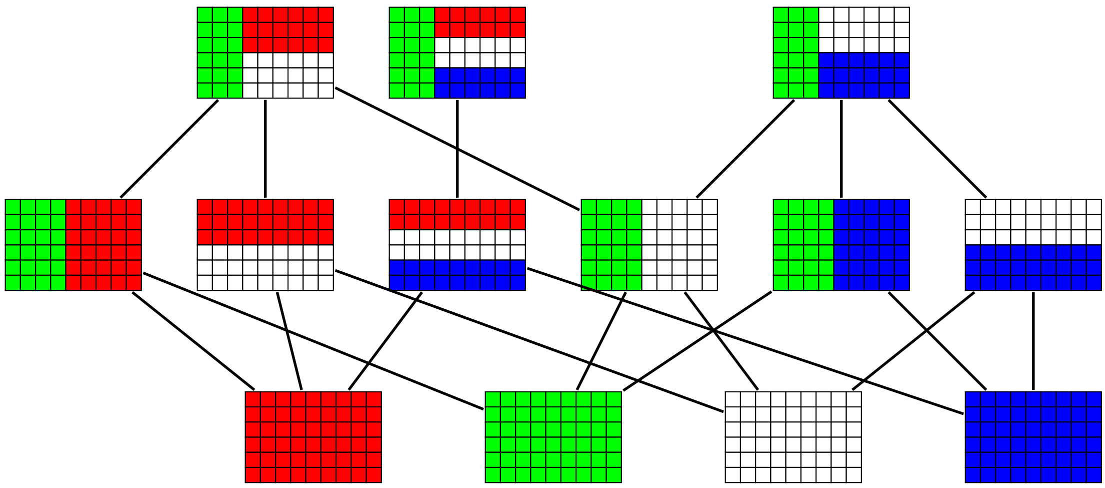


Figure C.1: Allowed patterns



An edge represents a relation $Q \prec P$ if P is above Q. Transitivity edges are not depicted. As an example $K \prec E$ and $K \prec C$.

Figure C.2: Hasse diagram of the order $\prec$ with the tile-set defined in Fig. C.1

Throughout the following, it will be more convenient for us to define tile-sets by the set of their *forbidden patterns*: a tile-set is then given by a finite set $\mathcal{F}_\tau$ of forbidden patterns ($\mathcal{F}_\tau = Q^V \setminus \mathcal{P}_\tau$); a configuration is a tiling if no forbidden pattern appears.

Let us now introduce the following natural preorder, which will play a central role in our paper:

Definition C.3 (*Preorder*)

Let x, y be two tilings, we say that $x \preceq y$ if any pattern that appears in x also appears in y.

We say that two tilings x, y are equivalent if $x \preceq y$ and $y \preceq x$. We denote this relation by $x \approx y$. In this case, x and y contain the same patterns. The equivalence class of x is denoted by $\langle x \rangle$. We write $x \prec y$ if $x \preceq y$ and $x \not\approx y$.

Some structural properties of tilings can be seen with the help of this preorder. The Hasse diagram in Fig. C.2 correspond to the relation $\prec$.

We choose to distinguish two types of tilings: A tiling x is of *type a* if any pattern that appears in x appears infinitely many times; x is of *type b* if there exists a pattern that appears only once in x. Note that any tiling is either of *type a* or of *type b*: suppose that there is a pattern that appears only a finite number of times in x; then the pattern which is the union of those patterns appears only once.

If x is of *type b*, then the only tilings equivalent to x are its shifted: there is a unique way in $\langle x \rangle$ to grow around the unique pattern.

C.2.2 Topology

In the domain of symbolic dynamics, topology provides both interesting results and is also a nice condensed way to express some combinatorial proofs [Hed69, GH55]. The benefit of topology is a little more surprising for tilings since they are essentially static objects. Nevertheless, we can get nice results with topology as will be seen in the sequel.

We see the space of configurations $Q^{\mathbb{Z}^2}$ as a metric space in the following way: the distance between two configurations c and c' is 2^{-i} where i is the minimal offset (for e.g., the euclidean norm) of a point where c and c' differ:

$$d(c, c') = 2^{-\min\{|i|,\ c(i) \neq c'(i)\}}$$

We could also endow Q with the discrete topology and then $Q^{\mathbb{Z}^2}$ with the product topology, thus obtaining the same topology as the one induced by d.

In this topology, a basis of open sets is given through the patterns: for each pattern P, the set $\mathcal{O}_P$ of all configurations c which contains P in their center (*i.e.* such that c is equal to P on its domain) is an open set, usually called a cylinder. Furthermore cylinders such defined are also closed (their complements are finite unions of $\mathcal{O}_{P'}$ where P' are patterns of same domain different from P). Thus $\mathcal{O}_P$'s are clopen.

Proposition C.1

$Q^{\mathbb{Z}^2}$ is a compact perfect metric space (a Cantor space).

We say that a set of configurations S is shift-invariant if any shifted version of any of its configurations is also in S; *i.e.* if for every $c \in S$, and every $v \in \mathbb{Z}^2$ the configuration c' defined by $c'(x) = c(x + v)$ is also in S. We denote such a shift by σ_v.

Remark. Our definition of pattern preorder C.3 can be reformulated in a topological way : $x \preceq y$ if and only if there exists shifts $(\sigma_i)_{i \in \mathbb{N}}$ such that $\sigma_i(y) \xrightarrow[i \to \infty]{} x$. We say that x can be extracted from y.

For a given configuration x, we define the topological closure of shifted forms of x: $\Gamma(x) = \overline{\{\sigma_v(x),\ v \in \mathbb{Z}^2\}}$ where $\sigma_{i,j}$ represents a shift of vector v.

We see that $x \preceq y$ if and only if $\Gamma(x) \subseteq \Gamma(y)$. Remark that x is minimal for $\prec$ if and only if $\langle x \rangle$ is closed.

As sets of tilings can be defined by a finite number of forbidden patterns, they correspond to *subshifts* of finite type [1]. In the sequel, we sometimes use arbitrary subshifts; they correspond to a set of configurations with a potentially infinite set of forbidden patterns.

1. *Subshifts* are closed shift-invariant subsets of $Q^{\mathbb{Z}^2}$

C.3 Main results

C.3.1 Basic structure

Let us first present a few structural results. First, the existence of minimal classes for $\prec$ is well known.

Theorem C.2 (*minimal elements*)

Every set of tilings contains a minimal class for $\prec$.

In the context of tilings, those that belong to minimal classes are often called *quasiperiodic*, while in language theory they are called *uniformly recurrent* or *almost periodic*. Those quasiperiodic configurations admit a nice characterization: any pattern that appears in one of them can be found in any sufficiently large pattern (placed anywhere in the configuration).

For a combinatorial proof of this theorem see [Dur99]. Alternatively, here is a scheme of a topological proof: consider a minimal subshift of $\mathcal{T}_\tau$ (such a subshift exists, see e.g., [RW92]) then every tiling in this set is in a minimal class.

An intensively studied class of tilings is the set of self-similar tilings. These tilings indeed are minimal elements (quasiperiodic) but one can find other kinds of minimal tilings (e.g., the nice approach of Kari and Culik in [II96]).

The existence of maximal classes of tilings is not trivial and we have to prove it:

Theorem C.3 (*maximal elements*)

Every set of tilings contains a maximal class for $\prec$.

Proof : Let us prove that any increasing chain has a least upper bound. The theorem is then obtained by Zorn's lemma.

Consider T_i an increasing chain of tiling classes. Consider the set P of all patterns that this chain contains. As the set of all patterns is countable, P is countable too, $P = \{p_i\}_{i\in\mathbb{N}}$.

Now consider two tilings T_k and T_l, any pattern that appears in T_k or T_l appears in $T_{\max(k,l)}$. Thus we can construct a sequence of patterns $(p'_i)_{i\in\mathbb{N}}$ such that p'_i contains all p_j, $j \leq i$ and p'_{i-1}. Note that p'_i is correctly tiled by the considered tile-set.

The sequence of patterns p'_i grows in size. By shift invariance, we can center each p'_i by superimposing an instance of p'_{i-1} found in p'_i over p'_{i-1}. We can conclude that this sequence has a limit and this limit is a tiling that contains all p_i, hence is an upper bound for the chain T_i. ■

Note that this proof also works when the *set of states* Q and/or the *set of forbidden patterns* $\mathcal{F}_\tau$ are countably infinite (neither compactness nor finiteness is assumed). However it is easy to construct examples where Q is infinite and there does not exist a minimal tiling.

Note that we actually prove that every chain has not only a upper bound, but also a least upper bound. Such a result does not hold for lower bound: We can easily build chains with lower bounds but no greatest lower bound.

C.3.2 Cantor-Bendixson

In this section we use the topological derivative and define Cantor-Bendixson rank; then we discuss properties of sets of tilings from this viewpoint. Most of the results presented in this section are direct translations of well known results in topology [Kur66].

A configuration c is said to be *isolated* in a set of configurations S if there exists a pattern P (of domain V) such that c is the only configuration in S that contains the pattern P in its center ($\forall x \in V, c(x) = P(x)$). We say that P *isolates* c. This corresponds to the topological notion: a point is isolated if there exists an open set that contains only this point. As an example, in Fig. C.3, the tilings A_i are isolated, the pattern isolating an A_i is the boundary between red, white, black and green parts of it.

The topological derivative of a set S is formed by its elements that are not isolated. We denote it by S'.

If S is a set of tilings, or more generally a subshift, we get some more properties. If P isolates a configuration in S then a shifted form of P isolates a shifted form of this configuration. Any configuration of S that contains P is isolated.

As a consequence, if $S = \mathcal{T}_\tau$, then $S' = \mathcal{T}_{\tau'}$ where τ' forbids the set $\mathcal{F}_\tau \cup \{P | P \text{ isolates some configuration in } \mathcal{T}_\tau\}$.

Note that S' is not always a set of tilings, but remains a subshift. Let us examine the example shown in Fig. C.3. S' is S minus the classes A_i. However any set of tilings (subshift of finite type) that contains C, B_i and D also contains A_i. Hence S' is not of finite type in this example.

We define inductively $S^{(\lambda)}$ for any ordinal λ :
- $S^{(0)} = S$
- $S^{(\alpha+1)} = (S^{(\alpha)})'$
- $S^{(\lambda)} = \bigcap_{\alpha<\lambda} S^{(\alpha)}$ when λ is a limit ordinal.

Notice that there exists a countable ordinal λ such that $S^{(\lambda+1)} = S^{(\lambda)}$. Indeed, at each step of the induction, the set of forbidden patterns increases, and there is at most countably many patterns. We call the least such ordinal the *Cantor-Bendixson rank of* S [Kur66].

An element c is *of rank* λ in S if λ is the least ordinal such that $c \notin S^{(\lambda)}$. If no such λ exists, c is of infinite rank. For instance all strictly quasiperiodic configurations (quasiperiodic configurations that are not periodic) are of infinite rank. We write $\rho(x)$ the rank of x.

An example of what Cantor-Bendixson ranks look like is shown in Fig. C.3, the first row contains the tilings of rank 1, the second row the ones of rank 2 etc.

Ranked tilings have many interesting properties. First of all, as any $\mathcal{T}_\tau^{(\lambda)}$ is shift-invariant, a tiling has the same rank as its shifted forms.

Note that at each step of the inductive definition, the set of isolated points is at most countable (there are less isolated points than patterns). As a consequence, if all tilings are ranked, $\mathcal{T}_\tau$ is countable, as a countable union (the Cantor-Bendixson rank is countable) of countable sets.

The converse is also true:

Theorem C.4

$\mathcal{T}_\tau$ is countable if and only if all tilings are ranked.

Proof : Let λ be the Cantor-Bendixson rank of $\mathcal{T}_\tau$. $\mathcal{T}_\tau^{(\lambda)} = \mathcal{T}_\tau^{(\lambda+1)}$ is a perfect set (no points are isolated). As a consequence, $\mathcal{T}_\tau^{(\lambda)}$ must be either empty or uncountable (classical application of Baire's Theorem : $\mathcal{T}_\tau^{(\lambda)}$ is compact thus has the Baire property and a non empty perfect set with the Baire property cannot be countable).

As $\mathcal{T}_\tau$ is countable, $\mathcal{T}_\tau^{(\lambda)} = \emptyset$. ■

Remark. Strictly quasiperiodic tilings only appear when the number of possible tilings is uncountable [Dur99]. As a consequence, if all tilings are ranked, strictly quasiperiodic tilings do not appear, thus all minimal tilings are periodic. In this case we therefore may expect all tilings to be somehow simple. We'll study this case later in this paper.

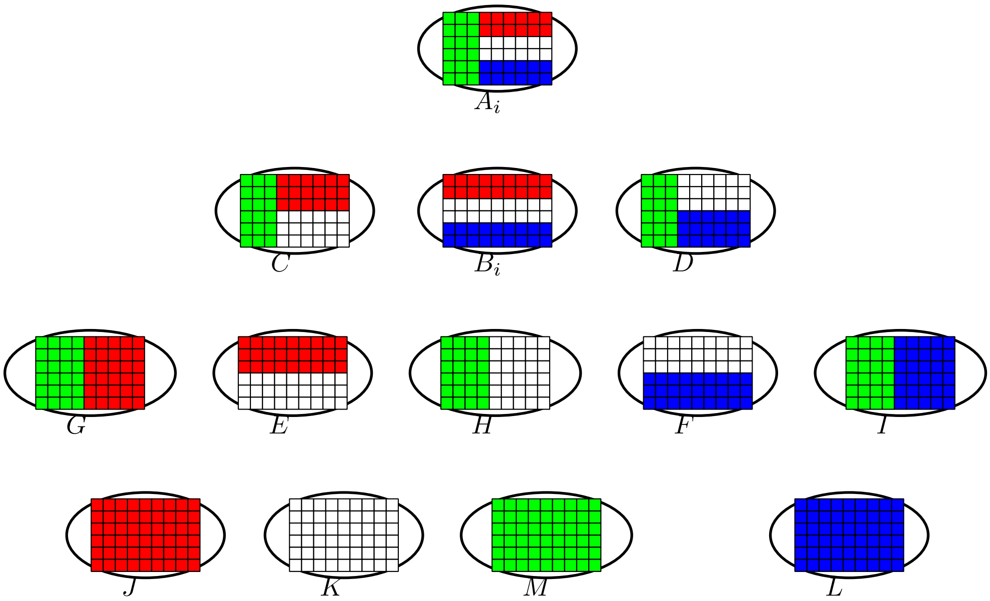


Figure C.3: Cantor-Bendixson ranks

As the topology of $Q^{\mathbb{Z}^2}$ has a basis of clopens $\mathcal{O}_P$, $Q^{\mathbb{Z}^2}$ is a 0-dimensional space, thus any subset of $Q^{\mathbb{Z}^2}$ is also 0-dimensional. As any (non empty) perfect 0-dimensional compact metric space is isomorphic to the Cantor Space we obtain:

Theorem C.5 (*Cardinality of tiling spaces*)

A set of tilings is either finite, countable or has the cardinality of continuum.

Note that the proof of this result does not make use of the continuum hypothesis.

We now present the connection between our preorder $\prec$ and the Cantor-Bendixson rank.

Proposition C.6

Let x and y be two ranked tilings such that $x \prec y$. Then $\rho(x) > \rho(y)$.

Proof : By definition of $\prec$, any pattern that appears in x also appears in y. As a consequence, if P isolates x in $S^{(\lambda)}$, then x is the only tiling of $S^{(\lambda)}$ that contains P hence y cannot be in $S^{(\lambda)}$. ■

Thus tilings of Cantor-Bendixson rank 1 (minimal rank) are maximal tilings for $\prec$. Conversely if all tilings are ranked, tilings of maximal rank exist and are minimal tilings. These tilings are periodic, see remark C.3.2.

Another consequence is that if all tilings are ranked, no infinite increasing chain for $\prec$ exists because such chain would induce an infinite decreasing chain of ordinals:

Theorem C.7

If $\mathcal{T}_\tau$ is countable, there is no infinite increasing chain for $\prec$.

C.3.3 The countable case

In the context of Cantor-Bendixson ranks, the case of countable tilings was revealed as an important particular case. Let us study this case in more details.

If the number of tilings is finite, the situation is easy: any tiling is periodic. Our aim is to prove that in the countable case, there exists a tiling c which has exactly one vector of periodicity (such a tiling is sometimes called weakly periodic in the literature).

We split the proof in three steps :

- There exists a tiling which is not minimal;
- There exists a tiling c which is at level 1, that is such that all tilings less than c are minimal;
- Such a tiling has exactly one vector of periodicity.

The first step is a result of independent interest. To prove the last two steps we use Cantor-Bendixson ranks.

Recall that in our case any minimal tiling is periodic (no strictly quasiperiodic tiling appears in a countable setting [Dur99]). The first step of the proof may thus be reformulated:

Theorem C.8

If all tilings produced by a tile-set are periodic, then there are only finitely many of them.

It is important to note that a compactness argument is not sufficient to prove this theorem, there is no particular reason for a converging sequence of periodic tilings with strictly increasing period to converge towards a non periodic tiling: there indeed exist such sequences with a periodic limit.

Proof : *We are in debt to an anonymous referee who simplified our original proof.*

Suppose that a tile-set produces infinitely many tilings, but only periodic ones.

As the set of tilings is infinite and compact, one of them is obtained as a limit of the others: There exists a tiling X and a sequence X_i of distinct tilings such that $X_i \to X$.

Now by assumption X is periodic of period p for some p. We may suppose that no X_i has p as a period. Denote by M the pattern which is repeated periodically.

$X_i \to X$ means that X_i contains in its center a square of size $q(i) \times q(i)$ of copies of M, where q is a growing function.

For each i, consider the largest square of X_i consisting only of copies of M. Such a largest square exists, as it is bounded by a period of X_i. Let k be the size of this square. Now, the boundary of this square contains a $p \times p$ pattern which is not M (otherwise this is not the largest square).

By shifting X_i so that this pattern is at the center, we obtain a tiling Y_i which contains a $p \times p$ pattern at the origin which is not M adjacent to a $k/2 \times k/2$ square consisting of copies of M in one of the four quarter planes.

By taking a suitable limit of these Y_i, we will obtain a tiling which contains a $p \times p$ pattern which is not M in its center adjacent to a quarter plane of copies of M.

Such a tiling cannot be periodic. ■

This proof does not assume that the set of forbidden patterns $\mathcal{F}_\tau$ is finite, therefore it is still valid for any shift-invariant closed subset (subshift) of $Q^{\mathbb{Z}^2}$.

Now we prove stronger results about the Cantor-Bendixson rank of $\mathcal{T}_\tau$. Let α be the Cantor-Bendixson rank of $\mathcal{T}_\tau$. Since $(\mathcal{T}_\tau)^{(\alpha)} = \emptyset$, α cannot be a limit ordinal: Suppose that it is indeed a limit ordinal, therefore $\bigcap_{\beta<\alpha}(\mathcal{T}_\tau)^{(\beta)} = \emptyset$ is an empty intersection of closed sets in $Q^{\mathbb{Z}^2}$ therefore by compactness there exists $\gamma < \alpha$ such that $\bigcap_{\beta<\gamma}(\mathcal{T}_\tau)^{(\beta)} = \emptyset$ and therefore $\mathcal{T}_\tau$ can not have rank α. Hence α is a successor ordinal, $\alpha = \beta + 1$.

However, we can refine this result :

Lemma C.9

The rank of $\mathcal{T}_\tau$ cannot be the successor of a limit ordinal.

Proof : Suppose that $\beta = \cup_{i<\omega}\beta_i$. Since $(\mathcal{T}_\tau)^{(\beta+1)} = \emptyset$, $(\mathcal{T}_\tau)^{(\beta)}$ is finite (otherwise it would have a non-isolated point by compactness), it contains only periodic tilings.

Let p be the least common multiple of the periods of the tilings in $(\mathcal{T}_\tau)^{(\beta)}$. Let M be the set of patterns of size $2p \times 2p$ that do not admit p as a period. Let x_i be an element that is isolated in $(\mathcal{T}_\tau)^{(\beta_i)}$.

As there is only a finite number of p-periodic tilings, we may suppose w.l.o.g. that no x_i admit p as a period.

For any i, there exists a pattern of M that appears in x_i. Let x'_i be the tiling with this pattern at its center. By compactness, one can extract a limit x' of the sequence $(x'_i)_{i\in\mathbb{N}}$, x' is by construction in $\cap_i(\mathcal{T}_\tau)^{(\beta_i)} = \mathcal{T}_\tau^{(\beta)}$. However, x' does not contain a $p-$periodic pattern at its center, that is a contradiction. ■

We write $\alpha = \lambda + 2$ the rank of $\mathcal{T}_\tau$.

We already proved that there exists a non minimal tiling but this is not sufficient to conclude that there exists a tiling at level 1 [2]. However, we achieve this as a corollary of the previous lemma: $(\mathcal{T}_\tau)^{(\lambda)}$ is infinite (otherwise $(\mathcal{T}_\tau)^{(\lambda+1)}$ would be empty) and contains a non periodic tiling by theorem C.8. This non periodic tiling c is not minimal (otherwise it would be strictly quasiperiodic and then $\mathcal{T}_\tau$ would not be countable). Now c is at level 1 : any tiling less than c is in $(\mathcal{T}_\tau)^{(\lambda+1)}$ therefore periodic (hence minimal).

If a tiling x is of *type a* and is ranked, then it has a vector of periodicity: consider the pattern P that isolates it in the last topological derivative of $\mathcal{T}_\tau$ that it belongs to. Since x is of *type a*, this pattern appears twice in it, therefore there exists a shift σ such that $\sigma(x)$ contains P at its center. $x = \sigma(x)$ because P isolates x.

As any tiling of *type a* has a vector of periodicity, it remains to prove that c is of *type a*:

Lemma C.10

c is of *type a*.

Proof : Suppose the converse : there exists a pattern P that appears only once in c. Considering the union of this pattern P and a pattern that isolates c, we may assume that P isolates c. c has only a finite number of tilings smaller than itself: they lie in $\mathcal{T}_\tau^{(\lambda+1)}$ which is finite, and are all periodic, say of period p. As P isolates c, none of these tilings contain P.

Consider the patterns of size $2p \times 2p$ of T that are not $p-$periodic. If those patterns can appear arbitrary far from P then one can extract a tiling from c (thus smaller than c) that is not $p-$periodic and does not contain P; this is not possible.

Therefore there is a pattern in c that contains P (thus appears only once) and any other part of c is $p-$periodic (one can gather all non $p-$periodic parts of c around P), as depicted in Fig. C.4a.

2. We actually can prove that the level 1 exists: There is no infinite decreasing chain whose lower bound is a periodic configuration

(a) What we get : c is periodic everywhere but at P

(b) P can appear at many different places since c has periodic patterns

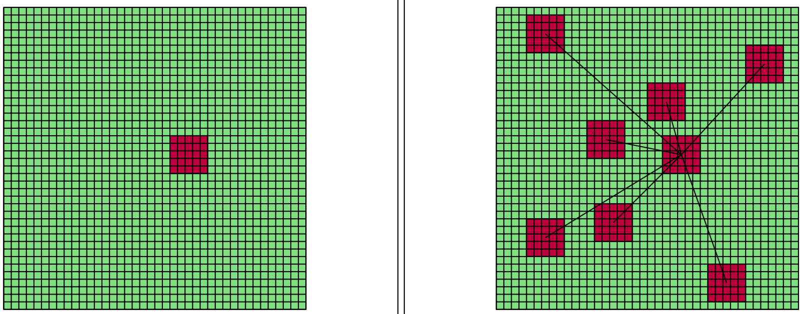


Figure C.4: What can happen if c is of *type b*?

This non periodic part could also be inserted at infinitely many different positions in c since the tiling rules are of bounded radius, as depicted in Fig. C.4b. Hence the number of tilings is not countable. ■

c is of *type a*, c is not periodic, c has a vector of periodicity, therefore our theorem C.11 holds:

Theorem C.11

If τ is a tile-set that produces a countable number of tilings then it produces a tiling with exactly one vector of periodicity.

C.4 Open problems

We are interested in proving more precise results for the order $\prec$ for a countable set of tilings: we wonder whether the order $\prec$ has at most finitely many levels, as it is the case in Fig. C.2. We know how to construct a tile-set so that the maximal level is any arbitrary integer see e.g., Fig.C.5 for level 3.

We also intend to prove a similar result for uncountable sets of tilings; the problem is that we are tempted to think that if the set of tilings is uncountable, then a quasiperiodic tiling must appear. However, this is not true: imagine a tile-set that admits a vertical line of white or black cells with red on the left and green on the right. The uncountable part is due to the vertical line that itself contains a quasiperiodic of dimension 1 but not of dimension 2. This tile-set produces tilings that looks like H in Fig. C.2, except that the vertical line can have two different colors without any constraint.

A generalization of lemma C.9 would be to prove that the Cantor-Bendixson rank of a countable set of tilings cannot be infinite; we know how to construct sets of tilings that have an arbitrary large but finite Cantor-Bendixson rank, but we do not know how to obtain a set of tilings of rank greater than ω.

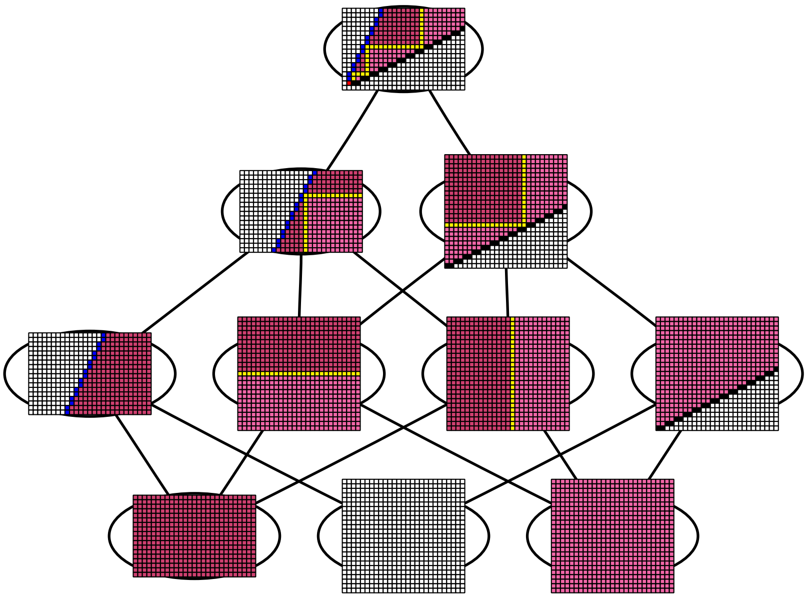


Figure C.5: An example of a tile-set that produces countably many tilings and a tiling at level 3

Annexe D

COMPUTING (OR NOT) QUASI-PERIODICITY FUNCTIONS OF TILINGS

L'article [I3], écrit avec Alexis Ballier, cherche à déterminiser si la fonction de quasipériodicité d'un pavage est toujours calculable. Dans le contexte de ce mémoire, on retiendra en particulier la construction présentée en D.3 qui construit un espace de pavages effectif où tous les minimaux ont un langage Turing-complet.

Abstract

We know that tilesets that can tile the plane always admit a quasi-periodic tiling [Bir12, Dur99], yet they hold many uncomputable properties [Ber64, Han74, Mye74, Sim11b]. The quasi-periodicity function is one way to measure the regularity of a quasi-periodic tiling. We prove that the tilings by a tileset that admits only quasi-periodic tilings have a recursively (and uniformly) bounded quasi-periodicity function. This corrects an error from [CD04, theorem 9] which stated the contrary. Instead we construct a tileset for which any quasi-periodic tiling has a quasi-periodicity function that cannot be recursively bounded. We provide such a construction for $1-$dimensional effective subshifts and obtain as a corollary the result for tilings of the plane *via* recent links between these objects [AS, DRS10].

Introduction

Tilings of the discrete plane as studied nowadays have been introduced by Wang in order to study the decidability of a subclass of first order logic [Wan60, Wan61, BGG01]. After Berger proved the undecidability of the domino problem [Ber64], interest has grown for understanding how complex are these simply defined objects [Han74, Mye74, DLS08, CD04]. Despite being able to have complex tilings, any tileset that can tile the plane admits a *quasi-periodic* tiling [Bir12, Dur99]; roughly speaking, a quasi-periodic tiling is a tiling in which every finite pattern can be found in any sufficiently large part of the tiling. It is therefore natural to define the quasi-periodicity function of a quasi-periodic tiling: it associates to an integer n the minimal size in which we are certain to find any pattern of size n [Dur99, CD04]. This is one way to measure the complexity of a quasi-periodic tiling and, to some extent, of a tileset τ since τ must admit at least one quasi-periodic tiling. We start by proving in Section D.2 that tilings by

1. Both authors are partly supported by ANR-09-BLAN-0164. A. Ballier has been partly supported by the Academy of Finland project 131558. We thank Pierre Guillon for discussions that lead to the constructions provided in Section D.3.

tilesets that admit only quasi-periodic tilings have a recursively (and uniformly) bounded quasi-periodicity function (Theorem D.2). Remark that there exists non-trivial tilesets that admit only quasi-periodic tilings [Rob71, Moz89, Oll08] and that the property of having only such tilings can be reduced to the domino problem [Ber64, Rob71] and is thus undecidable [2].

With the aim to study discretization of dynamical systems, $1-$dimensional subshifts have been extensively studied in symbolic dynamics [MH38, LM95]. Quasi-periodic tilings correspond to almost periodic sequences [Hed69] or uniformly recurrent sequences in this context. Again, the existence of complex uniformly recurrent sequences has been shown [MSU03]. In Section D.3 we show, given a partial recursive function φ, how to construct an effective subshift in which every uniformly recurrent configuration has a quasi-periodicity function greater than φ where it is defined (Theorem D.8). This allows us to correct the error from [CD04] as we obtain as a corollary (using recent links between tilings and effective $1-$dimensional subshifts [AS, DRS10]) that there exists tilesets for which no quasi-periodic tiling can have a quasi-periodicity function that is recursively bounded (Theorem D.10).

D.1 Definitions

A *configuration* is an element of $\mathbf{Q}^{\mathbb{Z}^2}$ where $\mathbf{Q}$ is a finite set or, equivalently, a mapping from $\mathbb{Z}^2$ to $\mathbf{Q}$. A *pattern* P is a function from a finite domain $\mathcal{D}_P \subseteq \mathbb{Z}^2$ to $\mathbf{Q}$. The *shift* of vector v ($v \in \mathbb{Z}^2$) is the function denoted by σ_v from $\mathbf{Q}^{\mathbb{Z}^2}$ to $\mathbf{Q}^{\mathbb{Z}^2}$ defined by $\sigma_v(c)(x) = c(v + x)$. A pattern P *appears* in a configuration c (denoted $\mathsf{P} \in c$) if there exists $v \in \mathbb{Z}^2$ such that $\sigma_v(c)_{|\mathcal{D}_P} = \mathsf{P}$. Similarly, we can define the shift of vector v of a pattern P by the function $\sigma_v(\mathsf{P})(x) = \mathsf{P}(v + x)$; then we can say that a pattern P appears in another pattern M if there exists $v \in \mathbb{Z}^2$ such that $\sigma_v(\mathsf{M})_{|\mathcal{D}_P} = \mathsf{P}$ and denote it by $\mathsf{P} \in \mathsf{M}$. We use the same vocabulary and notations for both notions of shift and appearance but there should not be any confusion since configurations are always denoted by lower case letters and patterns by upper case letters.

Given a finite set of colors $\mathbf{Q}$, a *tileset* is defined by a finite set of patterns $\mathcal{F}$; we say that a configuration c is a *valid tiling* for $\mathcal{F}$ if none of the patterns of $\mathcal{F}$ appear in c. We denote by $\mathbf{T}_\mathcal{F}$ the set of valid tilings for $\mathcal{F}$. If $\mathbf{T}_\mathcal{F}$ is non-empty we say that $\mathcal{F}$ can tile the plane. A set of configurations $\mathcal{S}T$ is said to be a *set of tilings* if there exists some finite set of patterns $\mathcal{F}$ such that $\mathcal{S}T = \mathbf{T}_\mathcal{F}$. This notion of set of tilings corresponds to subshifts of finite type [LM95, Lin04]. When we impose no restriction on $\mathcal{F}$ these are subshifts and when $\mathcal{F}$ is recursively enumerable we say that $\mathbf{T}_\mathcal{F}$ is an *effective subshift* (see, e.g., [DKB04, Hoc09b, Hoc09a, AS, DRS10]).

A periodic configuration c is a configuration such that the set $\left\{\sigma_v(c), v \in \mathbb{Z}^2\right\}$ is finite. It is well known (since Berger [Ber64]) that there exists tilesets that do not admit a periodic tiling but can still tile the plane. On the other hand, quasi-periodicity is the correct regularity notion if we always want a tiling with this property. Periodic configurations are quasi-periodic but the converse is not true. Several characterizations of quasi-periodic configurations exist [Dur99], we give one here that we use for the rest of the paper.

2. Take a tileset τ_u that admits only one uniform tiling (and thus only quasi-periodic tilings), a tileset τ_f that admits non quasi-periodic tilings (e.g., a fullshift on $\{0, 1\}$) then it is clear that $\left(\tau \times \tau_f\right) \cup \tau_u$ admits only quasi-periodic tilings if and only if τ does not tile the plane.

Definition D.1 (*Quasi-periodic configuration*)

A configuration $c \in \mathbf{Q}^{\mathbb{Z}^2}$ is quasi-periodic if any pattern that appears in c appears in any sufficiently large pattern of c.

More formally, if a pattern P appears in c then there exists $n \in \mathbb{N}$ such that for every pattern M defined on $[-n; n]^2$ that appears in c, P appears in M.

We denote by $n_{(\mathsf{P},c)}$ the smallest such n for finding a pattern P in the quasi-periodic configuration c.

Theorem D.1 (*[Bir12, Dur99]*)

Any non-empty set of tilings contains a quasi-periodic configuration.

For an integer n, the set of patterns defined on a square domain $[-n; n]^2$ is finite, it is therefore natural to define the quasi-periodicity function of a quasi-periodic configuration.

Definition D.2 (*Quasi-periodicity function*)

The quasi-periodicity function of a quasi-periodic configuration c, denoted by $\mathcal{Q}_c$, is the function from $\mathbb{N}$ to $\mathbb{N}$ that maps a given integer n to the smallest integer m such that any pattern of domain $[-n; n]^2$ that appears in c appears in any pattern of c of domain $[-m; m]^2$.

$$\mathcal{Q}_c(n) = \max \left\{ n_{(\mathsf{P},c)}, \mathsf{P} \in c, \mathcal{D}_P = [-n; n]^2 \right\}$$

The function $\mathcal{Q}_c$ measures in some sense the complexity of the quasi-periodic configuration c: the faster it grows, the more complex c is. Since one can construct tilesets whose tilings have many uncomputable properties (e.g., such that every tiling is uncomputable as a function from $\mathbb{Z}^2$ to $\mathbf{Q}$ [Han74, Mye74] or such that every pattern that appears in a tiling has maximal Kolmogorov complexity [DLS08]), it is natural to expect the quasi-periodicity function to inherit the non-recursive properties of tilings. This is what had been proved in [CD04].

In some particular cases it is easy to prove that this function is actually computable. Consider a tileset such that any pattern that appears in a tiling appears in every tiling; in that case every tiling is quasi-periodic and the quasi-periodicity function is the same for every tiling. Moreover there exists an algorithm that decides if a pattern can appear in a tiling or not (this has been proven by different ways, either by considering the fact that the first order theory of the tileset is finitely axiomatizable and complete therefore decidable [I8] or by using a direct compactness argument [Hoc09b]). Given this algorithm, it is easy to compute the quasi-periodicity function (that does not depend on the tiling): for a given p, compute all the $[-p; p]^2$ patterns that appear in a tiling and then compute all the $[-n; n]^2$ patterns for $n \geq p$ until every $[-p; p]^2$ pattern appears in every $[-n; n]^2$ pattern and output the smallest such n.

In the remainder of this paper, we improve this technique to obtain a less restrictive condition on the tileset while proving that the quasi-periodicity function is recursively bounded:

Theorem D.2

If a tileset (defined by $\mathcal{F}$) admits only quasi-periodic tilings then there exists a computable function $q : \mathbb{N} \to \mathbb{N}$ such that for any tiling c of $\mathbf{T}_{\mathcal{F}}$, c has a quasi-periodicity function bounded by q, *i.e.* $\forall c \in \mathbf{T}_{\mathcal{F}}, \forall n \in \mathbb{N}, \mathcal{Q}_c(n) \leq q(n)$.

Note that this result is contrary to a result in [CD04] stating that there exists tilesets admitting only quasi-periodic tilings with quasi-periodicity functions with no computable upper bound. There is indeed a mistake in [CD04] that will be examined later.

D.2 Computable bound on the quasi-periodicity function

In this section we consider a tileset defined by a finite set of forbidden patterns $\mathcal{F}$ such that every tiling by $\mathcal{F}$ is quasi-periodic. The only hypothesis we have is the following: For any tiling $c \in \mathbf{T}_{\mathcal{F}}$ and for any pattern P that appears in c, there exists an integer $n_{(\mathsf{P},c)}$ such that any $[-n_{(\mathsf{P},c)}, n_{(\mathsf{P},c)}]^2$ pattern that appears in c contains P. In order to prove Theorem D.2, we first have to prove that there exists a bound that does not depend on the tiling:

Lemma D.3

If a tileset $\mathcal{F}$ admits only quasi-periodic tilings then, for any pattern P that appears in some tiling of $\mathbf{T}_{\mathcal{F}}$, there exists an integer n such that any tiling that contains P also contains P in all its $[-n; n]^2$ patterns.

We define $n_{(\mathsf{P},\mathcal{F})}$ to be the smallest integer with this property.

Remark that the converse of this lemma is obviously true by definition: if for any pattern there exists such an integer then all the tilings are quasi-periodic.

Proof : Suppose this is not true: there exists a pattern P and a sequence $(c_n)_{n\in\mathbb{N}}$ of configurations that contain P and such that c_n also contains a $[-n; n]^2$ pattern that does not contain P.

For a given n, consider O_n, one of the largest square patterns of c_n that does not contain P. Since c_n is quasi-periodic and contains P by hypothesis, there does not exist arbitrary large square patterns that do not contain P and thus O_n is well defined. Note that O_n is defined on at least $[-n; n]^2$. Since we supposed O_n of maximal size, there must be a pattern P adjacent to it like depicted on Figure D.1.

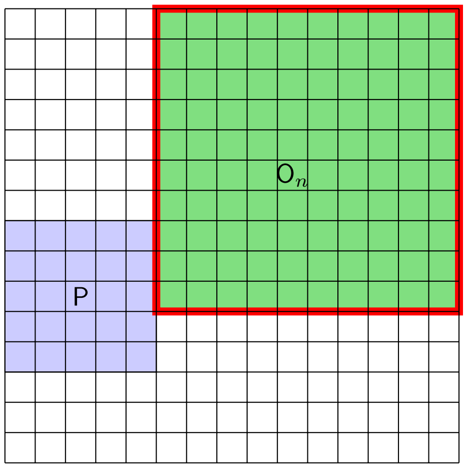


Figure D.1: O_n near P.

Now if we center our view on this P adjacent to O_n, for infinitely many n's the largest part of O_n always appears in the same quarter of plane (with origin P). Since O_n is defined on at least $[-n; n]^2$, by compactness we obtain a tiling with P at its center and a quarter of plane without P. Such a tiling cannot be quasi-periodic. ■

Lemma D.3 shows that if all the tilings that are valid for $\mathcal{F}$ are quasi-periodic then there must exist a global bound on the quasi-periodicity function of any tiling. Indeed we define

$f(n) = \max\left\{n_{(\mathsf{P},\mathcal{F})}, \mathcal{D}_P = [-n;n]^2, \mathsf{P} \text{ appears in a tiling by } \mathcal{F}\right\}$; for any tiling $c \in \mathbf{T}_{\mathcal{F}}$ and any integer n, we have $\mathcal{Q}_c(n) \leq f(n)$. The only part left in the proof of Theorem D.2 is to prove that f is computably bounded.

In a quasi-periodic tiling, if a pattern P defined on $[-n;n]^2$ appears in it then it must appear close to P (at distance less than $f(n)+n$) in each of the four quarters of plane starting from the corners of P. In general, we cannot compute whether a pattern will appear in some tiling or not, however, we can compute whether a pattern is valid with respect to $\mathcal{F}$.

Lemma D.4

If a tileset $\mathcal{F}$ admits only quasi-periodic tilings then, for any pattern P defined on $[-n;n]^2$ that appears in some tiling of $\mathbf{T}_{\mathcal{F}}$, there exists an integer m such that any pattern R defined on $[-n-m;n+m]^2$ that is valid with respect to $\mathcal{F}$ and contains P at its center (*i.e.* $\mathsf{R}_{|[-n;n]^2} = \mathsf{P}$) is such that the four patterns $\mathsf{R}_{|[-n-m;-n]^2}$, $\mathsf{R}_{|[-n-m;-n]\times[n;n+m]}$, $\mathsf{R}_{|[n;n+m]\times[-n-m;-n]}$, $\mathsf{R}_{|[n;n+m]^2}$ all contain P.

We define $m_{(\mathsf{P},\mathcal{F})}$ to be the smallest integer m with this property.

Those four patterns may seem obscure at a first read, they are depicted on Figure D.2.

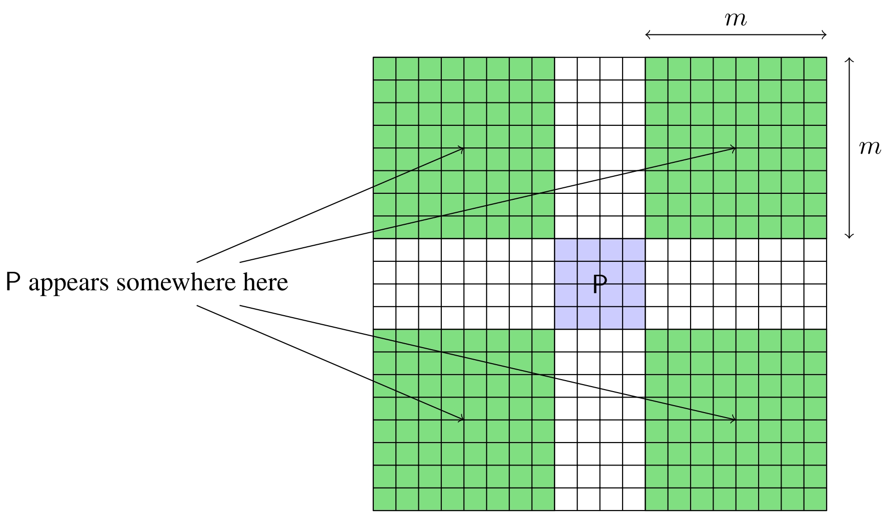


Figure D.2: The four patterns in which we must find another occurrence of P.

Proof : For a given pattern P, suppose that there exists no such m. This means that there exists arbitrary large m and valid patterns $\mathsf{R_m}$ (defined on $[-n-m;n+m]^2$) such that one of the four patterns $\mathsf{R_m}_{|[-n-m;-n]^2}$, $\mathsf{R_m}_{|[-n-m;-n]\times[n;n+m]}$, $\mathsf{R_m}_{|[n;n+m]\times[-n-m;-n]}$, $\mathsf{R_m}_{|[n;n+m]^2}$ does not contain P.

Without loss of generality, we can assume that this always happens in the same quarter of plane. By extracting a tiling centered on the pattern P at the center of $\mathsf{R_m}$ (which we can do by compactness), there exists a tiling c of $\mathbf{T}_{\mathcal{F}}$ that contains P and a quarter of plane without P, contradicting the quasi-periodicity of c. ■

Note that the converse of Lemma D.4 is also true: if, for any pattern P, there exists such an

$m_{(\mathsf{P},\mathcal{F})}$ then all the tilings of $\mathbf{T}_{\mathcal{F}}$ are quasi-periodic.

Lemma D.5

If $\mathcal{F}$ is a tileset that allows only quasi-periodic tilings then, for any pattern P defined on $[-p;p]^2$ that appears in some tiling of $\mathbf{T}_{\mathcal{F}}$, we have:

$$n_{(\mathsf{P},\mathcal{F})} \leq 2(m_{(\mathsf{P},\mathcal{F})} + p)$$

Proof : Let c be a (quasi-periodic) tiling of $\mathbf{T}_{\mathcal{F}}$ that contains P and a pattern O defined on $[-k;k]^2$ that does not contain P with $k > 2(m_{(\mathsf{P},\mathcal{F})} + p)$. Without loss of generality, we may assume that O is of maximal size. That is, there is a pattern P adjacent to O. Let R be the pattern defined on $[-p - m_{(\mathsf{P},\mathcal{F})}; p + m_{(\mathsf{P},\mathcal{F})}]^2$ centered on the pattern P adjacent to O in c. Since $k > 2(m_{(\mathsf{P},\mathcal{F})} + p)$ and O does not contain P, at least one of the four patterns $\mathsf{R}_{|[-p-m_{(\mathsf{P},\mathcal{F})};-p]^2}$, $\mathsf{R}_{|[-p-m_{(\mathsf{P},\mathcal{F})};-p]\times[p;p+m_{(\mathsf{P},\mathcal{F})}]}$, $\mathsf{R}_{|[p;p+m_{(\mathsf{P},\mathcal{F})}]\times[-p-m_{(\mathsf{P},\mathcal{F})};-p]}$, $\mathsf{R}_{|[p;p+m_{(\mathsf{P},\mathcal{F})}]^2}$ does not contain P as depicted on Figure D.3; since R is a valid pattern with respect to $\mathcal{F}$, this contradicts the definition of $m_{(\mathsf{P},\mathcal{F})}$. ■

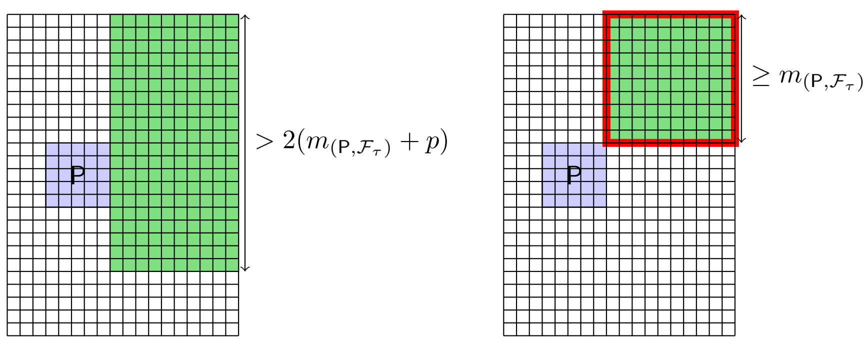


Figure D.3: Bounding the size of the patterns not containing P.

Now that we have a bound that deals only about locally valid patterns instead of patterns that appear in tilings (and therefore is computably checkable), we can proceed to the proof of Theorem D.2:

Proof (of Theorem D.2): $\mathcal{F}$ is a tileset that admits only quasi-periodic tilings. For an integer n, compute all the patterns $\mathsf{P}_1, \ldots, \mathsf{P}_\mathsf{k}$ defined on $[-n;n]^2$ that are valid for $\mathcal{F}$.

For each of these P_j use the following algorithm: For each integer i, compute the set $\mathsf{R}_1, \ldots, \mathsf{R}_\mathsf{p}$ of patterns defined on $[-i-n; i+n]^2$ that contain P_j at their center and are valid with respect to $\mathcal{F}$.

1. If there is no such pattern R, claim that P_j cannot appear in any tiling by $\mathcal{F}$, and define e.g., $b_{\mathsf{P}_\mathsf{j}} = 0$. Then continue with $\mathsf{P}_{\mathsf{j+1}}$
2. If all these patterns R restricted to either $[-n-i;-n]^2$, $[-n-i;-n] \times [n;n+i]$, $[n;n+i] \times [-n-i;n]$ or $[n;n+i]^2$ all contain P then define $b_{\mathsf{P}_\mathsf{j}} = 2(i+n)$ and continue with $\mathsf{P}_{\mathsf{j+1}}$ [3].

3. Remark that these patterns are exactly those depicted in Figure D.2.

For any pattern, one of these cases always happens: If $\mathsf{P_j}$ appears in at least one tiling of $\mathbf{T}_{\mathcal{F}}$ then, by Lemma D.4, for $i = m_{(\mathsf{P_j},\mathcal{F})}$ we are in case 2. If $\mathsf{P_j}$ does not appear in any tiling of $\mathbf{T}_{\mathcal{F}}$ then case 1 must happen, otherwise we would have arbitrary large extensions of $\mathsf{P_j}$ and hence a tiling containing $\mathsf{P_j}$ by compactness. Note that we may halt in case 2 even if $\mathsf{P_j}$ does not appear in any tiling.

Now compute $q(n) = \max\left\{b_{\mathsf{P_j}}, \mathcal{D}_{P_j} = [-n;n]^2\right\}$.

For any tiling $c \in \mathbf{T}_{\mathcal{F}}$ and any pattern P defined on $[-n;n]^2$ that appears in c we have:

$$\begin{array}{rcll} n_{(\mathsf{P},c)} & \leq & n_{(\mathsf{P},\mathcal{F})} & \text{by definition of } n_{(\mathsf{P},\mathcal{F})} \\ & \leq & 2(m_{(\mathsf{P},\mathcal{F})} + n) & \text{by Lemma D.5} \\ & \leq & b_{\mathsf{P}} & \text{by minimality of } m_{(\mathsf{P},\mathcal{F})} \\ & \leq & q(n) & \text{by definition of } q \end{array}$$

Therefore, for any configuration c and any integer n, we have $\mathcal{Q}_c(n) \leq q(n)$ and q is the computable function that completes the proof of Theorem D.2. ■

We remark that all the arguments used in the proofs of the lemmas involve only compactness and the fact that we can decide if a given pattern is valid for $\mathcal{F}$. Hence, we may remove some restrictions on $\mathcal{F}$: $\mathbf{T}_{\mathcal{F}}$ is still compact if $\mathcal{F}$ is infinite and we can still decide if a given pattern is valid for $\mathcal{F}$ when $\mathcal{F}$ is recursive. Moreover, if $\mathcal{F}$ is recursively enumerable then there exists a recursive set of patterns $\mathcal{F}'$ such that $\mathbf{T}_{\mathcal{F}} = \mathbf{T}_{\mathcal{F}'}$: consider the (computable) enumeration $f(0), f(1), \ldots$ of $\mathcal{F}$; when enumerating $f(i)$, we can compute an integer n such that all the previously enumerated patterns are defined on a domain included in $[-n;n]^2$; then we enumerate all the extensions of $f(i)$ defined on $[-n-1;n+1]^2 \cup \mathcal{D}_{f(i)}$. This enumeration enumerates a new set of patterns $\mathcal{F}'$ that is now recursive since they are enumerated by increasing sizes. It is straightforward that $\mathbf{T}_{\mathcal{F}} = \mathbf{T}_{\mathcal{F}'}$. We conclude that $\mathcal{F}$ needs not to be finite in order for Theorem D.2 to be valid but we may assume that it is only recursively enumerable. Sets of tilings with a recursively enumerable set of forbidden patterns are usually called *effective subshifts* in the literature [DKB04, Hoc09b, Hoc09a, AS, DRS10] or also Π^0_1 subshifts [Sim11b, Mil11] and are a special case of effectively closed sets as studied in computable analysis (see e.g., [Wei00]) [4].

D.3 Large quasi-periodicity functions

In this section we prove that we can construct tilesets such that all its quasi-periodic tilings have a large quasi-periodicity function. We start from a 1-dimensional effective subshift $\mathcal{S}X$ over an alphabet Σ and then build an effective subshift over the alphabet $\Sigma \times \{0,1\}$, and the complexity of the quasi-periodicity function will come from the top layer. For this, consider all occurrences of a word u in the subshift $\mathcal{S}X$. There are infinitely many of them, so the top layer restricted to occurrences of u will contain a bi-infinite word over $\{0,1\}$. If we can find infinitely many words in the subshift $\mathcal{S}X$ so that occurrences of different words do not somehow overlap in a configuration c, then this would give us an infinite number of bi-infinite words within a single configuration c, in which we could code something.

The following lemma tells us how to find such words in the general case of minimal effective subshifts; a *minimal subshift* is a subshift in which every pattern that appears in a configuration

4. The definitions are usually given in dimension one, *i.e.* for (bi-)infinite, words even though they are the same for multi-dimensional configurations.

appears in every configuration, or equivalently, a subshift that does not admit a proper non-empty subshift. In this case, all configurations are of course quasi-periodic.

Lemma D.6

For any (non-empty) $1-$dimensional minimal effective subshift $\mathcal{S}X \subseteq \Sigma^{\mathbb{Z}}$ that has no periodic configuration there exists a computable sequence $(u_n)_{n\in\mathbb{N}}$ of words in the language of $\mathcal{S}X$ such that no u_n is prefix of another one.

Proof : We build recursively a sequence $(u_0, \ldots, u_n)$ and a word v_n such that the set $\{u_k, k \leq n\} \cup \{v_n\}$ is prefix-free. For $n = 0$, take two different letters in Σ ($|\Sigma| > 1$ comes from the hypothesis as $\mathcal{S}X$ is non-empty and does not contain any periodic configuration).

Now suppose we obtain $(u_0, \ldots u_n)$ and v_n. Since $\mathcal{S}X$ is supposed to be minimal, v appears in an uniformly recurrent way in a configuration of $\mathcal{S}X$ and since $\mathcal{S}X$ contains no periodic configuration, there exists two different right-extensions of v: w and w' of the same length. Taking $u_{n+1} = w$ and $v_{n+1} = w'$ ends the recurrence. ■

To obtain our theorem, we will need a subshift $\mathcal{S}X$ for which we control precisely the sequence u_n.

Lemma D.7

There exists a (non-empty) 1-dimensional minimal effective subshift $\mathcal{S}X$ and a computable sequence $(u_n)_{n\in\mathbb{N}}$ of words in the language of $\mathcal{S}X$ so that $|u_n| \leq n$ and no u_n is prefix of another one.

Proof : We will use a construction based on Toeplitz words. Let p be an integer. For an integer n, denote by $\phi_p(n)$ the first non-zero digit in the writing of n in base p, e.g., $\phi_3(15) = 2$.

Let $w_p = \phi_p(1)\phi_p(2)\ldots$. For example $w_4 = 123112321233123112311 23\ldots$.

Now let $\mathcal{S}X_p$ be the shift of all configurations c so that all words of c are words of w_p. Note that any word of size n appearing in w_p appears at a position less than p^n so that $\mathcal{S}X_p$ is an effective subshift.

Now the following statements are clear:

- For every word w in w_p, there exists k so that for every configuration $c \in \mathcal{S}X_p$, w appears periodically in c of period p^k (w might appear in some other places)
- $\mathcal{S}X_p$ is minimal (a consequence of the previous statement)

If u_1 and u_2 are two words over Σ_1 and Σ_2 of the same size, we write $u_1 \otimes u_2$ for the word over $\Sigma_1 \times \Sigma_2$ whose ith projection is u_i ($i \in \{1, 2\}$).

Now let $\mathcal{S}X = \mathcal{S}X_7 \otimes \mathcal{S}X_8$. $\mathcal{S}X$ is a shift, and $\mathcal{S}X$ is minimal[5]: If $c_1 \otimes c_2 \in \mathcal{S}X_7 \otimes \mathcal{S}X_8$ and u_1 and u_2 are two patterns resp. of w_7 and w_8 of the same size, then u_1 appears periodically in c_1 of period 7^{k_1} and u_2 appears periodically in c_2 of period 8^{k_2}. As these two numbers are relatively prime, there exists a common position i so that u_1 (resp. u_2) appears in position i in c_1 (resp c_2), so that $u_1 \otimes u_2$ appears in $c_1 \otimes c_2$.

Now we can find the sequence u_n.

Let u be a word in $\{5, 6, 7\}^\star\{1, 2, 3, 4\}$. We define $f_8(u)$ inductively as follows:

- If $|u| = 1$, then $f_8(u) = u$.
- If $u = xu_1$ then let $v = f_8(u_1)$ and n be the length of v.

5. Note that the Cartesian product of two minimal shifts is not always minimal [Sal10].

– If $x = 5$ then $f_8(u) = 567v_1 1234567v_2 1234567v_3 1 \dots v_n 1234$
– If $x = 6$ then $f_8(u) = 67v_1 1234567v_2 1234567v_3 1 \dots v_n 12345$
– If $x = 7$ then $f_8(u) = 7v_1 1234567v_2 1234567v_3 1 \dots v_n 123456$

Now it is clear that each $f_8(u)$ is in w_8 and by a straightforward induction, no $f_8(u)$ is prefix of another. Let $S_8 = \{f_8(u) | u \in \{5,6,7\}^\star \{1,2,3,4\}\}$ Note that $f_8(u)$ is of length $8^{|u|-1}$. In particular we have $4 \times 3^{n-1}$ words of length 8^{n-1} in S_8.

We do the same with w_7, with words $u \in \{4,5,6\}^\star\{1,2,3\}$, to obtain a set S_7 containing $3 \times 3^{n-1}$ words of length 7^{n-1}. We can always enlarge all words in S_7 to obtain a set S'_7 containing $3 \times 3^{n-1}$ words of length 8^{n-1}.

Now take $S = S'_7 \otimes S_8$. This set contains $12 \times 9^{n-1} > 8^n$ words of size 8^{n-1} for each n and no word of S is prefix of one another. Now an enumeration in increasing order of S gives the sequence $(u_n)_{n\in\mathbb{N}}$.

The whole construction is clearly effective. ∎

Theorem D.8

Given a partial computable function φ, there exists a $1-$dimensional effective subshift $\mathcal{S}X_\varphi$ such that any quasi-periodic configuration c in $\mathcal{S}X_\varphi$ has a quasi-periodicity function $\mathcal{Q}_c$ such that $\mathcal{Q}_c(n) \geq \varphi(n)$ when $\varphi(n)$ is defined.

Proof : Consider the subshift $\mathcal{S}X$ and the computable sequence $(u_n)_{n\in\mathbb{N}}$ that are given by Lemma D.7. Since Lemma D.7 ensures that $|u_n| \leq n$, a sequence $(u_n)_{n\in\mathbb{N}}$ with the additional property that $|u_n| = n$ is also computable since we can compute an extension of the words u_n in $\mathcal{S}X$ since it is minimal and effective and the prefix-free property is retained while taking extensions. We assume this additional property in this proof.

Let $\Sigma' = \Sigma \times \{0,1\}$. We define $\mathcal{S}X_\varphi$ as a subshift of $\mathcal{S}X \times \{0,1\}^{\mathbb{Z}}$.

Compute in parallel all the $\varphi(n)$. When $\varphi(n)$ is computed we add the following additional constraints: On the $\{0,1\}$ layer of Σ' we force a 1 to appear on the first letter of u_n once every $\varphi(n)+1$ occurrences of u_n, the first letter of all other occurrences of u_n being 0. There is no ambiguity since no u_n is prefix of another one. This defines $\mathcal{S}X_\varphi$ as an effective subshift since $\mathcal{S}X$ is effective and $(u_n)_{n\in\mathbb{N}}$ is computable.

Every u_n appears in every configuration of $\mathcal{S}X$ since it is minimal. If $\varphi(n)$ is defined, then every u_n with a 1 on the $\{0,1\}$ layer appears exactly every $\varphi(n)$ occurrences of u_n's with a 0 on its $\{0,1\}$ layer in every configuration of $\mathcal{S}X_\varphi$. Therefore, for any quasi-periodic configuration c of $\mathcal{S}X_\varphi$ we have that $\mathcal{Q}_c(n) \geq \varphi(n)$ where $\varphi(n)$ is defined which completes the proof. ∎

Corollary D.9

There exists a 1-dimensional effective subshift $\mathcal{S}X$ such that every quasi-periodic configuration c in $\mathcal{S}X$ has a quasi-periodicity function which is not bounded by any computable function.

Proof : Let $(\varphi_n)_{n\in\mathbb{N}}$ be an effective enumeration of partial computable functions.

Let $\varphi(n) = \varphi_n(n) + 1$; φ is also a partial computable function; we can therefore find an effective one dimensional subshift $\mathcal{S}X_\varphi \subseteq \Sigma^{\mathbb{Z}}$ via Theorem D.8 such that any quasi-periodic configuration c of $\mathcal{S}X_\varphi$ is such that $\mathcal{Q}_c \geq \varphi$ where φ is defined, hence $\mathcal{Q}_c$ is not recursively bounded. ∎

Theorem D.10

There exists a tileset such that every quasi-periodic tiling has a quasi-periodicity function that is not recursively bounded.

Proof : Take the effective $1-$dimensional subshift of the previous corollary (as a subshift of $\Sigma^{\mathbb{Z}}$): $\mathcal{S}X_{\varphi}$. There exists a set of tilings (or $2-$dimensional SFT) $\mathcal{S}X_{\varphi}^{2} \subseteq (Q \times \Sigma)^{\mathbb{Z}^2}$ encoding it [AS, DRS10] in the following way:

In any configuration of $\mathcal{S}X_{\varphi}^{2}$, the rows of the $\Sigma-$layer are identical, that is, if we write this configuration as $c_Q \times c_\Sigma \in Q^{\mathbb{Z}^2} \times \Sigma^{\mathbb{Z}^2}$, for any i, j in $\mathbb{Z}$, $c_\Sigma(i,j) = c_\Sigma(i,j+1)$. Moreover, the projection:

$$\begin{array}{rccl} p: & (Q \times \Sigma)^{\mathbb{Z}^2} & \to & \Sigma^{\mathbb{Z}} \\ & c_Q \times c_\Sigma & \to & \begin{array}{ccc} \mathbb{Z} & \to & \Sigma \\ n & \to & c_\Sigma(n,0) \end{array} \end{array}$$

of $\mathcal{S}X_{\varphi}^{2}$ is exactly $\mathcal{S}X_{\varphi}$ (*i.e.* $p(\mathcal{S}X_{\varphi}^{2}) = \mathcal{S}X_{\varphi}$). Since the configurations of $\mathcal{S}X_{\varphi}^{2}$ are the Cartesian product of a construction layer (the $Q^{\mathbb{Z}^2}$ part) and the effective $1-$dimensional subshift $\mathcal{S}X_{\varphi}$ repeated on the rows, the quasi-periodicity function of any quasi-periodic configuration of $\mathcal{S}X_{\varphi}^{2}$ is greater or equal to the quasi-periodicity function of the quasi-periodic 1-dimensional configuration it represents. ■

Note that quasi-periodicity configurations obtained in the constructions in [AS, DRS10] are rather benign. If we start from a $1-$dimensional quasi-periodic configuration c, then the quasi-periodic tilings x that are projected onto c have a quasi-periodicity function that is computable knowing the quasi-periodicity function of c.

D.4 Note

Theorem 9 in [CD04] stated the contrary of Theorem D.2: “there exists a tileset such that all its tilings are quasi-periodic and none of its quasi-periodicity function is computably bounded”. Besides some errors that can be easily corrected, there is a big problem in the construction they claim to give. They encode K, a recursively enumerable but not recursive set, in every tiling in a way such that if $i \in K$ then it must appear in every tiling in a pattern of size $g(i)$ where g is a computable function. This property allows by itself to decide K: For an integer i, compute $g(i)$ and all the possible encodings of i if it were to appear in a tiling; patterns that do not appear in a tiling of the plane are recursively enumerable [6] and thus, when we have enumerated all the patterns coding i we know that $i \notin K$. Since K is supposed recursively enumerable, this allows to decide K.

6. Simply try to tile arbitrary big patterns around it and if it is not possible claim that the pattern does not appear in a tiling.

Annexe E

PERIODICITY IN TILINGS

Cet article [I4] écrit avec Pascal Vanier montre que l'ensemble des périodes possibles d'un TSFT ne peut pas être quelconque : Au contraire, il est lié à la classe de complexité ***NP****. A noter que les idées présentes dans cet article sont également utilisées dans [I5, A1].*

Abstract

Tilings and tiling systems are an abstract concept that arise both as a computational model and as a dynamical system. In this paper, we prove an analog of the theorems of Fagin [Fag74] and Selman and Jones [JS72b] by characterizing sets of periods of tiling systems by complexity classes.

Introduction

The model of tilings was introduced by Wang [Wan61] in the 60s to study decision problems on some classes of logical formulas. Roughly speaking, we are given some local constraints (a *tiling system*), and we consider colorings of the plane that respect these constraints (a *tiling*).

Tilings are both a simple and powerful model: the definitions are quite easy to grasp, however most decision problems on tilings are computationally intractable, starting from the most important one: decide whether a given tiling system tiles the plane [Ber66]. This is due in part to a straightforward encoding of Turing machines in tilings (see e.g. [Büc62a, vEB97]) and to the existence of *aperiodic* tiling systems, i.e. that produce many tilings, but none of them being periodic. In fact, Harel [Har85] and van Emde Boas [vEB97] give strong evidence that tilings are more suitable than Turing machines and satisfiability problems to express hardness in both complexity and recursivity theory.

In a similar manner, we will show in this paper how to do *descriptive complexity* [EF95] with tilings. More precisely, we will prove that sets of *periods* of tilings correspond exactly to non-deterministic exponential time, hence to spectra of first order formulas [JS72b]. As a consequence, the problems whether non-deterministic exponential time, first order spectra, or sets of periods are closed under complementation are equivalent.

The result is in itself not surprising: in fact, the proof of Jones and Selman [JS72b] on spectra of first order formulas uses a multi-dimensional generalization of finite automata which may be interpreted as a kind of tiling system. Furthermore, the encoding of Turing machines in tilings, in particular in periodic tilings, is well known. This paper makes use of the fact that n steps of a Turing machine can be encoded into a tiling of period exactly n. Usual proofs [AD01, GK72] have a quadratic blowup, which is unsuitable for our purpose.

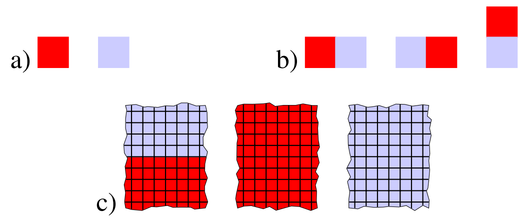


Figure E.1: The set of tiles (a) and the forbidden patterns (b) can only form tilings of the forms shown in (c).

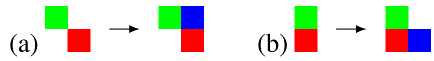


Figure E.2: North-East (NE) determinism (a) and East (E) determinism (b) in a tiling system

E.1 Tilings, periodicity and computations

For any dimension $d \geq 1$, a tiling of $\mathbb{Z}^d$ with a finite set of tiles T is a mapping $c : \mathbb{Z}^d \to T$. A $\mathbb{Z}^d$-tiling system is the pair (T, F), where F is a finite set of forbidden patterns $F \subset T^N$, where N is a finite subset of $\mathbb{Z}^d$ called the neighborhood. A tiling c is said to be *valid* if and only if none of the patterns of F ever appear in c. Since the number of forbidden patterns is finite, we could specify the rules by *allowed* patterns as well. We give an example of such a tiling system with the tiles of figure E.1a and the forbidden patterns of figure E.1b. The allowed tilings are shown in figure E.1c.

The *cartesian product* of two tiling systems τ_1 and τ_2 is the tiling system obtained by superimposing the tiles of τ_1 to the tiles of τ_2 with the rules of τ_1 on the layer of τ_1 and the rules of τ_2 on the layer of τ_2. We will sometimes restrict the resulting tiling system by removing some superimposition of tiles.

A tiling c of dimension $d = 2$ is said to be *horizontally periodic* if and only if there exists a *period* $p \in \mathbb{N}^\star$ such that for all $x, y \in \mathbb{Z}$, $c(x, y) = c(x + p, y)$. A tiling c of $\mathbb{Z}^d$ is *periodic* if it has the same period on all its dimensions:

$$c(x_1, x_2, \ldots, x_d) = c(x_1 + p, x_2, \ldots, x_d) = \cdots = c(x_1, x_2, \ldots, x_d + p)$$

The smallest such p is called the *(horizontal) eigenperiod* of c.

A tiling system is *aperiodic* if and only if it tiles the plane but there is no valid periodic [1] tiling. Such tiling systems have been shown to exist [Ber66] and are at the core of the undecidability of the *domino problem* (decides if a given tiling system tiles the plane). J. Kari [Kar92, KP99] gave such a tiling system with an interesting property: determinism. A tiling system is *NE-deterministic* (for *North-East*) if given two tiles respectively at the southern and western neighbour of a given cell, there is at most one tile that can be put in this cell so that the finite pattern is valid. The mechanism is shown in figure E.2a.

It is easy to change some details in order to have an other form of determinism: two tiles vertically adjacent will force their lower right neighboring tile as shown in figure E.2b. This type of determinism will be called East-determinism. With such a tiling system, if a column of the plane is given, the half plane on its right is then determined.

1. On none of the dimensions.

As said earlier, tilings and recursivity are intimately linked. In fact, it is quite easy to encode Turing machines in tilings. Such encodings can be found e.g. in [Kar94, Cha08]. Given a Turing machine M, we can build a tiling system τ_M in figure E.3. The tiling system is given by *Wang tiles*, i.e., we can only glue two tiles together if they coincide on their common edge. This tiling system τ_M has the following property: there is an accepting path for the word u in time (less than) t using space (less than) w if and only if we can tile a rectangle of size $(w+2)\times t$ with white borders, the first row containing the input. Note that this method works for both deterministic and non-deterministic machines.

E.2 Recognizing languages with tilings

Let τ be a $\mathbb{Z}^d$-tiling system. Then we define

$$\mathcal{L}_\tau = \{n \mid \text{there exists a tiling by } \tau \text{ of eigenperiod } n\}$$

If τ is a $\mathbb{Z}^2$-tiling system, we define

$$\mathcal{L}^h_\tau = \{n \mid \text{there exists a tiling by } \tau \text{ of horizontal eigenperiod } n\}$$

Definition E.1

A set $L \subseteq \mathbb{N}^\star$ is in $\mathcal{T}$ if $L = \mathcal{L}_\tau$ for some $\mathbb{Z}^d$-tiling system τ and some d. A set $L \subseteq \mathbb{N}^\star$ is in $\mathcal{H}$ if $L = \mathcal{L}^h_\tau$ for some $\mathbb{Z}^2$-tiling system τ.

$\mathcal{T}$ and $\mathcal{H}$ are the classes of languages recognized by tiling and recognized horizontally by tiling respectively.

We say that a language $\mathcal{L}$ is recognized by a tiling system τ if and only if $\mathcal{L} = \mathcal{L}^h_\tau$ or $\mathcal{L} = \mathcal{L}_\tau$, depending on the context. It is easy to see that $\mathcal{T}$ and $\mathcal{H}$ are closed under union. It is not clear whether they are closed under intersection: if τ and τ' are tiling systems, a natural way to do intersection is to consider the *cartesian product* of τ and τ'. However, if for example $\mathcal{L}_\tau = \{2\}$ and $\mathcal{L}_{\tau'} = \{3\}$, then there exists in τ a tiling of eigenperiod 2, hence of period 6, and the same is true for τ', so that in this example $\tau \times \tau'$ will contain a tiling of eigenperiod 6 whereas $\mathcal{L}_\tau \cap \mathcal{L}_{\tau'} = \emptyset$.

In this paper, we prove the following:

Theorem E.1

$\mathcal{H}$ is closed under union, intersection and complementation.

Theorem E.2

$\mathcal{T}$ is closed under intersection. $\mathcal{T}$ is closed under complementation if and only if **NE** $=$ **coNE**.

Here **NE** is the class of languages recognized by a (one-tape) non-deterministic Turing machine in time 2^{cn} for some $c > 0$. Note that for theorem E.2 we need to work in any dimension d. That is, if $\mathcal{L}_1, \mathcal{L}_2$ are the sets of periods of the $\mathbb{Z}^d$-tiling systems τ_1, τ_2, then there exists a $\mathbb{Z}^{d'}$-tiling system τ' that corresponds to the set of period $\mathcal{L}_1 \cap \mathcal{L}_2$. However d' can be larger than d.

To prove these theorems, we will actually give a characterization of our classes $\mathcal{H}$ and $\mathcal{T}$ in terms of structural complexity. This will be the purpose of the next two sections.

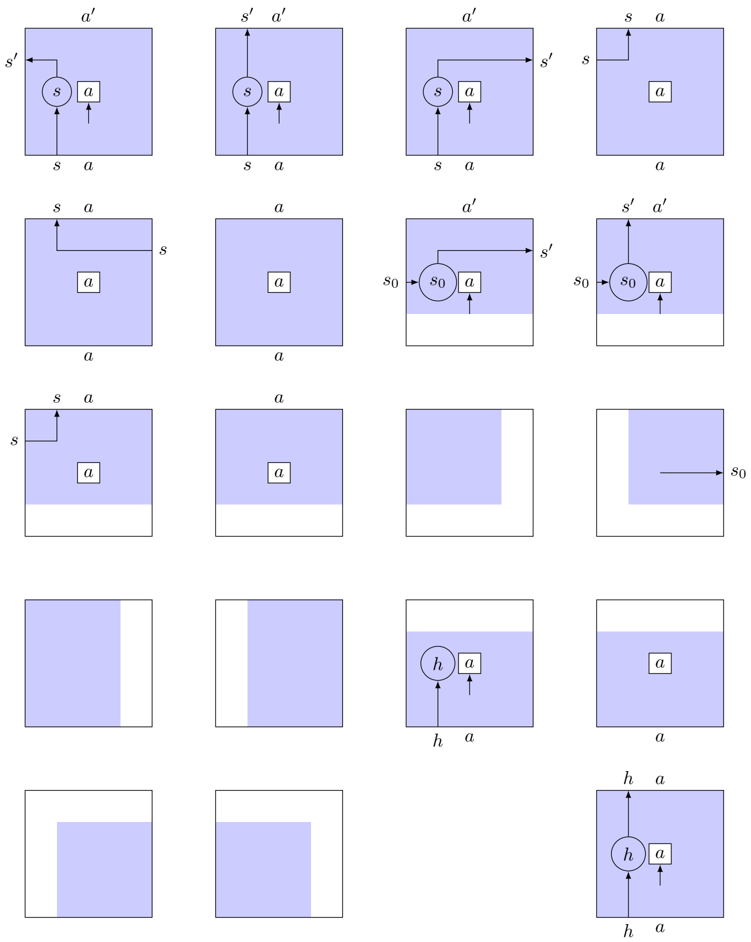


Figure E.3: A tiling system, given by Wang tiles, simulating a Turing machine. The meaning of the labels are the following: (a) label s_0 represents the initial state of the Turing machine. (b) The top-left tile corresponds to the case where the Turing machine, given the state s and the letter a on the tape, writes a', moves the head to the left and to change from state s to s'. The two other tiles are similar. (c) h represents a halting state. Note that the only states that can appear in the last step of a computation (before a border appears) are halting states.

E.3 $\mathcal{H}$ and NSPACE(2^n)

To formulate our theorem, we consider sets of periods, i.e. subsets of $\mathbb{N}^\star$, as unary or binary languages. If $L \subset \mathbb{N}^\star$ then we define $un(L) = \{1^{n-1} | n \in L\}$. We define $bin(L)$ to be the set of binary representations (missing the leading one) of numbers of L. As an example, if $L = \{1, 4, 9\}$, then $un(L) = \{\epsilon, 111, 11111111\}$ and $bin(L) = \{\epsilon, 00, 001\}$. Note that any language over the letter 1 (resp. the letters $\{0, 1\}$) is the unary (resp. binary) representation of some subset of $\mathbb{N}^\star$.

We now proceed to the statement of the theorem:

Theorem E.3

Let $\mathcal{L}$ be a language, the following statements are equivalent:

i) $\mathcal{L} \in \mathcal{H}$

ii) $un(\mathcal{L}) \in$ **NSPACE**(n)

iii) $bin(\mathcal{L}) \in$ **NSPACE**(2^n)

Recall that **NSPACE**(n) is the set of languages recognized by a (one-tape) non-deterministic Turing machine in space $\mathcal{O}(n)$.

The $(ii) \Leftrightarrow (iii)$ is folklore from computational complexity theory. The following two lemmas will prove the equivalence $(ii) \Leftrightarrow (i)$ hence the result.

Lemma E.4

For any tiling system τ, $un(\mathcal{L}^h_\tau) \in$ **NSPACE**(n).

Proof : Let $\tau = (T, F)$ be a tiling system. We will construct a non-deterministic Turing machine accepting 1^n if and only if $n + 1$ is a horizontal eigenperiod of τ. The machine has to work in space $\mathcal{O}(n)$, the input being given in unary.

Let $r > 0$ be an integer such that all patterns in the neighborhood N are smaller than a $r \times r$ square. Then a configuration c is correctly tiled if and only if all $r \times r$ blocks of c are correctly tiled. Furthermore, we can prove that if a horizontally periodic tiling of period n exists, then we can find such a tiling which is also vertically periodic of period at most $|T|^{rn}$. Now we give the algorithm, starting from n as an input:

- Initialize an array P of size n so that $P[i] = 1$ for all i.
- First choose non-deterministically $p \leq |T|^{rn}$
- Choose r bi-infinite rows $(c_i)_{0 \leq i \leq r-1}$ of period n (that is, choose $r \times n$ tiles).
- For each $r + 1 \leq i \leq p$, choose a bi-infinite row c_i of period n (that is, choose n tiles), and verify that all $r \times r$ blocks in the rows $c_i \ldots c_{i-r+1}$ are correctly tiled. At each time, keep only the last r rows in memory (we never forget the r first rows though).
- (Verification of the eigenperiod) If at any of the previous steps, the row c_i is not periodic of period $k < n$, then $P[k] := 0$
- For $i \leq r$, verify that all $r \times r$ blocks in the rows $c_{p-i} \ldots c_p c_0 \ldots c_{i-r-1}$ are correctly tiled
- If there is some k such that $P[k] = 1$, reject. Otherwise accept.

This algorithm needs to keep only $2r$ rows at each time in memory, hence is in space $\mathcal{O}(n)$. ■

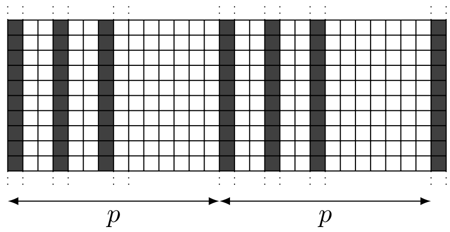


Figure E.4: A periodic tiling with the tiling system A.

Lemma E.5

For any unary language $L \in \mathbf{NSPACE}(n)$, then $\{n \in \mathbb{N}^\star \mid 1^{n-1} \in L\} \in \mathcal{H}$.

Proof : Let $L \in \mathbf{NSPACE}(n)$, there exists then a non-deterministic Turing machine M accepting L in linear space. Using traditional tricks from complexity theory, we can suppose that on input 1^n the Turing machine uses exactly $n+1$ cells of the tape (i.e. the input, with one additional cell on the right) and works in time exactly c^n for some constant c.

We will build a tiling system τ so that $1^n \in L$ if and only if $n+4$ is a period of the tiling τ. The modification to obtain $n+1$ rather that $n+4$, and thus prove the lemma, is left to the reader (basically “fatten” the gray tiles presented below so that they absorb 3 adjacent tiles), and serve no interest other than technical.

The proof may basically be split into two parts: First produce a tileset so that every tiling of horizontal period n looks like a grid of rectangles of size n by c^n delimited by gray cells (see fig. E.6b). Then encode the Turing machine M inside these rectangles. The main difficulty is in the first part, the second part being relatively straightforward. Note however, that as M is nondeterministic, the computation in different rectangles might be different. To not break the periodicity, we will have to *synchronize* all the machines.

The tiling system will be made of several components (or *layers*), each of them having a specific goal. The components and their rules are as follows:

– The first component A is composed of an aperiodic E-deterministic tiling system, whose tiles will be called "whites". We take the one from section E.1. We add a "gray" tile. The rules forbid any pattern containing a white tile above or below a gray tile. Hence a column containing a gray tile can only have gray tiles. We also forbid for technical reasons two gray tiles to appear next to each other horizontally.
With this construction, a periodic tiling of period p must have gray columns, as the white tiles form an aperiodic tileset.
For the moment nothing forbids more than one gray column to appear inside a period. Figure E.4 shows a possible form of a periodic tiling at this stage.
– The second component $D = P \times \{R, B\}$ will produce gray rows so that the (horizontally periodic) tiling will consist of $n \times c^n$ white rectangles delimited by these gray columns and rows.
The idea is as follows: suppose each word between two gray columns is a word over the alphabet $\{0, \ldots c-1\}$, that is, represents a number k between 0 and $c^n - 1$. Then it is easy with a tiling system to ensure that the number on the *next* line is $k+1$ (with the convention $(c^n - 1) + 1 = 0$). See figure E.3 for a transducer in the case $c = 2$ and its realization as a tiling system.

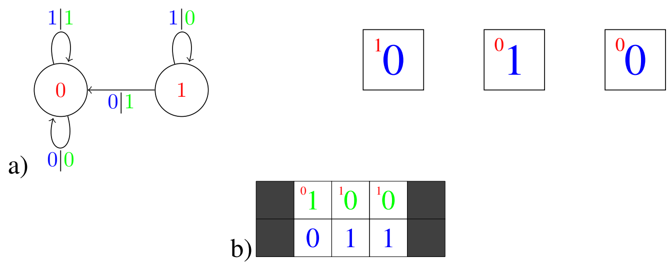


Figure E.5: a) The transducer doing the addition of one bit and the corresponding tiles. A valid tiling with these tiles is given in c).

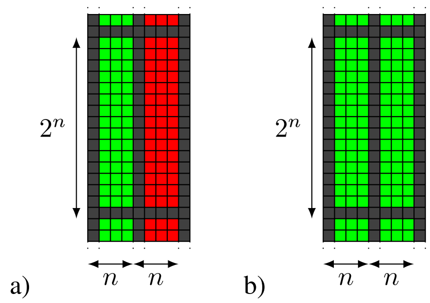


Figure E.6: a) The form after adding the component D, the aperiodic components between two gray columns can be different. b) After adding also the component T, the aperiodic components are exactly the same.

With the $\{R, B\}$ subcomponent, we mark the lines corresponding to the number 0, so that one line out of c^n is marked. We proceed as follows. First, horizontal bi-infinite lines are uniform: that is a tile is R if and only if its left neighbour is R. Next, a gray tile is R if and only if it is at the west of a 0-tile and at the north-west of a $c-1$-tile. That is, lines that are colored in R represent the time when the number is reinitialized from $c^n - 1$ to 0.

Figure E.6a shows some typical tiling at this stage: The period of a tiling is not necessarily the same as the distance between the rectangles, it may be larger. Indeed, the white tiles in two consecutive rectangles may be different.

– Component T is only a copy of A (without the same rules) which allows us to synchronize the first white columns after the gray columns: synchronising these columns ensures that the aperiodic components between two gray columns are always the same, since the aperiodic tiles are E-deterministic. The rules are simple, two horizontal neighbors have the same value on this component and a tile having a gray tile on its left has the same value in A as in T.

  At this stage, we have regular rectangles on all the plane, whose width correspond to the period of the tiling, as shown in figure E.6b.

– The last component M is the component allowing us to encode Turing machines in each rectangle. We use the encoding τ_M we described previously in section E.1. We force the computation to appear inside the white tiles: the white bottom borders must appear only in the row R, and the row below will have top borders. And the two tiles between the row R and the gray tiles are corner tiles. Finally, the input of the Turing

machine (hence the row above the row R) consists of only "1" symbols, with a final blank symbol.
The Turing machines considered here being non-deterministic, there could be different valid transitions on two horizontally adjacent rectangles, that is why we synchronize the transitions on each row. The method for the synchronization of the transition is almost the same as the method for the synchronisation of the aperiodic components, and thus not provided here.

Now we prove that $1^n \in L$ if and only if $n+4$ is a period of the tiling system τ. First suppose that n is a period and consider a tiling of period n.

- Due to component A, a gray column must appear. The period is a succession of either gray and white columns.
- Due to the component D, the gray columns are spaced by a period of p, $p < n$.
- Due to component T the tiling we obtain is (horizontally) p-periodic when restricted to the components D, T, A.
- For the component M to be correctly tiled, the input 1^{p-4} (4 = 1 (gray) + 1 (left border) + 1 (right border) + 1 (blank marker)) must be accepted by the Turing machine, hence $1^{p-4} \in L$
- Finally, due to the synchronization of the non-deterministic transition, the M component is also p-periodic. As a consequence, our tiling is p-periodic, hence $n = p - 4$. Therefore $1^{n+4} \in L$

Conversely, suppose $1^n \in L$. Consider the coloring of period $n+4$ obtained as follows (only a period is described):

- The component A consists of $n+3$ correctly tiled columns of our aperiodic E-deterministic tiling systems, with an additional gray column. As the E-deterministic tiling system tiles the plane, such a tiling is possible
- The component M corresponds to a successful computation path of the Turing machine on the input 1^n, that exists by hypothesis. As the computation lasts exactly than c^n steps, the computation fits exactly inside the $n \times c^n$ rectangle.
- We then add all other layers according to the rules to obtain a valid configuration, thus obtaining a valid tiling of period exactly $n+4$. ∎

Corollary E.6

The languages recognized horizontally by tiling are closed under intersection and complementation.

Proof : Immerman-Szelepczenyi's theorem [Imm88, BDG88] states that non-deterministic space complexity classes are closed under complementation. The result is then a consequence of theorem E.3. ∎

This theorem could be generalized to tilings of dimension d by considering tilings having a period n on $d-1$ dimensions, as explained in [Bor08].

E.4 $\mathcal{T}$ and NE

We now proceed to total periods rather than horizontal periods. We will prove:

Theorem E.7

Let $\mathcal{L} \subset \mathbb{N}^\star$ be a language, the following statements are equivalent:

i) There exists a $\mathbb{Z}^d$-tiling system τ for some d so that $\mathcal{L} = \mathcal{L}_\tau$

ii) $un(\mathcal{L}) \in$ **NP**

iii) $bin(\mathcal{L}) \in$ **NE**

In these cases, $\mathcal{L}$ is also the spectrum of a first order formula, see [JS72b].

We will obtain as a corollary theorem E.2. Note the slight difference in formulation between theorem E.7 and theorem E.3. While we can encode a Turing machine working in time n in a tiling of size n^2, we cannot check the validity of the tiling in less than $\mathcal{O}(n^2)$ time steps. More generally, it is unclear whether we can encode a Turing machine working in time n^c in a tiling of size less than n^{2c}. To overcome this gap, we need to work in any dimension d: A language $L \in$ **NTIME**(n^d) will be encoded into a tiling in dimension $2d$ and a tiling in dimension d will be encoded into a language L in **NTIME**(n^d).

The gap here is not surprising: while space complexity classes are usually model independent, this is not the case for time complexity, where the exact definition of the computational model matters. An exact characterization of periodic tilings for $d = 2$ is in fact possible, but messy: it would involve Turing machines working in space $O(n)$ with $O(n)$ reversals, see e.g [CM06].

Proof : The statements $(i) \Rightarrow (ii) \Leftrightarrow (iii)$ were already explained. So we only have to prove $(ii) \Rightarrow (i)$. We will see how a language $L \in$ **NTIME**(n^d) will be encoded as periods in dimension $2d$. The proof is similar to the previous one, and we provide only a sketch due to the lack of space. There are two steps:

– Build a tileset τ so that every tiling of period n looks like a lattice of hypercubes of size n delimited by gray cells.
– Encode the computation of the Turing Machine inside the cubes.

First step

We will work in dimension 2, as it is relatively straightforward to adapt the technique to higher dimensions.

The idea is to take an aperiodic NE-deterministic tiling system as white tiles. Then we add in this component three gray tiles: a cross tile, a horizontal tile and a vertical tile: gray horizontal (resp. vertical) lines consist of horizontal (resp. vertical) tiles, and they can intersect only on cross tiles.

Let A be this component. Now consider a periodic tiling. This tiling cannot contain only white tiles. Hence, it contains a horizontal tile, a vertical tile or a cross tile. The problem is that the presence of a horizontal(resp. vertical) tile does not imply the presence of a vertical (resp. horizontal) tile. The idea is to use in the next component an unary counter in *both* directions, so that it creates horizontal lines between vertical columns, and conversely.

The next components are also simple. The key point is that we use a NE-deterministic tiling system (rather than an E-deterministic) to ensure that all white squares contain the same tiles: in a NE-deterministic tiling, a square is entirely determined by its first row and

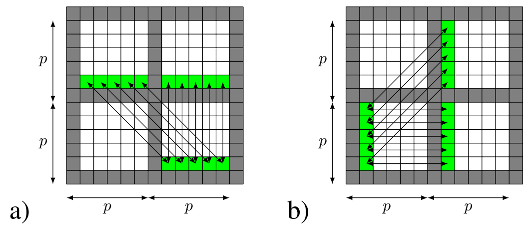


Figure E.7: Transmission of the first row in a) and of the first column in b).

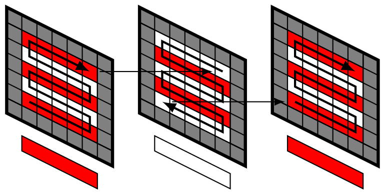


Figure E.8: Folding of a three dimensional cube, the red on the parity layer stands for +1 and the white for -1. The direction where to look for the next cell is given by the parity layer.

its first column. Hence it is sufficient to synchronize the first row and the first column of all squares to synchronize the aperiodic components. A way to do this is given in figure E.7.

Second step We now have to explain how to encode the computation of a Turing Machine working in time n^d into a cube of size n in dimension $2d$. The idea is to *fold* the space-time diagram of the Turing machine so that if fits into the cube. Each cell of the space-time diagram has coordinates (t, s) with $t \leq n^d, s \leq n^d$. We now have to transform each cell (t, s) into a cell of the hypercube of size n in dimension $2d$ so that two consecutive (in time or space) cells of the space-time diagram correspond to two adjacent cells of the cube, so that we can verify *locally* that the cube indeed encodes the computation of the Turing machine. This is exactly what a reflected n-ary Gray code [Flo56, Knu05] does.

Such a folding has already been described by Borchert [Bor08] and can also be deduced from [JS72b]. Basically, the cell at position $(t_0, \ldots, t_{d-1}) \in [0, n-1]^d$ will represent the integer $t = \sum a_i n^i$ where $a_i = t_i$ if $\sum_{j>i} t_j$ is even, and $a_i = n - 1 - t_i$ otherwise (formula (51) in [Knu05]). The noticeable fact is that the direction in which to look for the next positition is given by the parity of the sum of the *stronger* bits. Hence, we will encode these parity layers in the tiling, an example for a three dimensional folding can be seen on figure E.8.

However recall that Turing machines are non-deterministic, so we have to synchronize the transitions between the different hypercubes, as was done in the previous theorem. Similar techniques can be used. ■

Concluding remarks

The results presented in this paper establish a link between computational complexity and tilings, in particular between the complexity classes **NSPACE**(2^n), **NE** and the sets of periods of the tilings. The result is very different from the sets of periods we can obtain for 1-dimensional tilings (subshifts of finite type, see [LM95])

For 1-dimensional tilings, the number c_n of tilings of period n is enclosed in the *zeta* function: $\zeta(z) = \exp\left(\sum \frac{c_n}{n} z^n\right)$ The zeta function is well understood in dimension 1 [LM95]. A zeta function for tilings of the plane has been described by Lind [Lin96]. To be thorough, one needs to count not only tilings of period n as we did in this article, but tilings of period Γ, where Γ is any lattice. We think a complete characterization of sets of periods for general lattices Γ of $\mathbb{Z}^2$ is out of reach. If we accept to deal only with square periods (that is with the lattices $n\mathbb{Z} \times n\mathbb{Z}$), as we did here, preliminary work suggests we can characterize the number of periodic tilings via the class $\#$**E**.

Annexe F

TURING DEGREES OF MULTIDIMENSIONAL SFTs

Cet article écrit avec Pascal Vanier présente une construction prouvant que tout ensemble Π_1^0peut être réalisé par un TSFT qutite à ajouter des points récursifs. Les résultats contenus ici sont rappelés à la fois dans les chapitres 2 et 3. Il s'agit ici de la version longue de l'article et non pas de la version conférence [I1].

Abstract

In this paper we are interested in computability aspects of subshifts and in particular Turing degrees of 2-dimensional SFTs (i.e. tilings). To be more precise, we prove that given any Π_1^0subset P of $\{0,1\}^{\mathbb{N}}$ there is a SFT X such that $P \times \mathbb{Z}^2$ is recursively homeomorphic to $X \setminus U$ where U is a computable set of points. As a consequence, if P contains a recursive member, P and X have the exact same set of Turing degrees. On the other hand, we prove that if X contains only non-recursive members, some of its members always have different but comparable degrees. This gives a fairly complete study of Turing degrees of SFTs.

Introduction

Wang tiles have been introduced by Wang [Wan61] to study fragments of first order logic. Independently, subshifts of finite type (SFTs) were introduced to study dynamical systems. From a computational and dynamical perspective, SFTs and Wang tiles are equivalent, and most recursive-flavoured results about SFTs were proved in a Wang tile setting.

Knowing whether a tileset can tile the plane with a given tile at the origin (also known as the origin constrained domino problem) was proved undecidable by Wang [Wan63]. Knowing whether a tileset can tile the plane in the general case was proved undecidable by Berger [Ber64, Ber66].

Understanding how complex, in the sense of recursion theory, the points of an SFT can be is a question that was first studied by Myers [Mye74] in 1974. Building on the work of Hanf [Han74], he gave a tileset with no recursive tilings. Durand/Levin/Shen [DLS08] showed, 40 years later, how to build a tileset for which all tilings have high Kolmogorov complexity.

A Π_1^0class is an effectively closed subset of $\{0,1\}^{\mathbb{N}}$, or equivalently the set of oracles on which a given Turing machine does not halt. Π_1^0sets occur naturally in various areas in computer science and recursive mathematics, see e.g. [CR98, Sim11a] and the upcoming book [CR11]. It is easy to see that any SFT is a Π_1^0class (up to a recursive coding of $\Sigma^{\mathbb{Z}^2}$ into $\{0,1\}^{\mathbb{N}}$). This has various consequences. As an example, every non-empty SFT contains a point which is not Turing-hard (see Durand/Levin/Shen [DLS08] for a self-contained proof). The main question is how different SFTs are from Π_1^0classes. In the one-dimensional case, some answers to these questions were given by Cenzer/Dashti/King/Tosca/Wyman [Das08, CDK08, CDTW10].

The main result in this direction was obtained by Simpson [Sim11b], building on the work of Hanf and Myers: for every Π_1^0class S, there exists a SFT with the same *Medvedev* degree as S. The Medvedev degree roughly relates to the "easiest" Turing degree of S. What we are interested in is a stronger result: *can we find for every Π_1^0set S a SFT which has the same Turing degrees ?* We prove in this article that this is true if S contains a recursive point but not always when this is not the case. More exactly we build (Theorem F.6) for every Π_1^0class S a SFT for which the set of Turing degrees is exactly the same as for S with the additional Turing degree of recursive points. We also show that SFTs that do not contain any recursive point always have points with different but comparable degrees (Corollary F.20), a property that is not true for all Π_1^0classes. In particular there exists Π_1^0classes that do not have any points with comparable degrees.

As a consequence, as every *countable* Π_1^0class contains a recursive point, the question is solved for countable sets: the sets of Turing degrees of countable Π_1^0classes are the same as the sets of Turing degrees of countable sets of tilings. In particular, there exist countable sets of tilings with some non-recursive points. This can be thought as a two-dimensional version of Theorem 8 in [CDTW10].

This paper is organized as follows. After some preliminary definitions, we start with a quick proof of a generalization of Hanf, already implicit in Simpson [Sim11b]. We then build a very specific tileset, which forms a grid-like structure while having only countably many tilings, all of them recursive. This tileset will then serve as the main ingredient to prove the result on the case of classes with a recursive point in section F.4. In section F.5 we finally show the result on classes without recursive points.

F.1 Preliminaries

F.1.1 SFTs and tilings

We give here some standard definitions and facts about multidimensional subshifts, one may consult Lind [Lin04] for more details. Let Σ be a finite alphabet, the d-dimensional full shift on Σ is the set $\Sigma^{\mathbb{Z}^d} = \left\{c = (c_x)_{x\in\mathbb{Z}^d} \middle| \forall x \in \mathbb{Z}^d, c_x \in \Sigma\right\}$. For $v \in \mathbb{Z}^d$, the shift functions $\sigma_v : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$, are defined locally by $\sigma_v(c_x) = c_{x+v}$. The full shift equipped with the distance $d(x, y) = 2^{-\min\left\{\|v\| \middle| v\in\mathbb{Z}^d, x_v\neq y_v\right\}}$ is a compact, perfect, metric space on which the shift functions act as homeomorphisms. An element of $\Sigma^{\mathbb{Z}^d}$ is called a *configuration*.

Every closed shift-invariant (invariant by application of any σ_v) subset X of $\Sigma^{\mathbb{Z}^d}$ is called a *subshift*. An element of a subshift is called a point of this subshift.

Alternatively, subshifts can be defined with the help of forbidden patterns. A *pattern* is a function $p : P \to \Sigma$, where P is a finite subset of $\mathbb{Z}^d$. Let $\mathcal{F}$ be a collection of *forbidden* patterns, the subset X_F of $\Sigma^{\mathbb{Z}^d}$ containing only configurations having nowhere a pattern of F. More formally, $X_\mathcal{F}$ is defined by

$$X_\mathcal{F} = \left\{x \in \Sigma^{\mathbb{Z}^d} \middle| \forall z \in \mathbb{Z}^d, \forall p \in F, x_{z+P} \neq p\right\}.$$

In particular, a subshift is said to be a *subshift of finite type* (SFT) when the collection of forbidden patterns is finite. Usually, the patterns used are *blocks* or *n-blocks*, that is they are defined over a finite subset P of $\mathbb{Z}^d$ of the form $[\![0, n-1]\!]^d$.

Given a subshift X, a block or pattern p is said to be *extensible* if there exists $x \in X$ in which p appears, p is also said to be extensible to x.

In the rest of the paper, we will use the notation Σ_X for the alphabet of the subshift X.

A subshift $X \subseteq \Sigma_X^{\mathbb{Z}^2}$ is a *sofic shift* if and only if there exists a SFT $Y \subseteq \Sigma_Y^{\mathbb{Z}^2}$ and a local map $f : \Sigma_Y \to \Sigma_X$ such that for any point $x \in X$, there exists a point $y \in Y$ such that for all $z \in \mathbb{Z}^d, x_z = f(y_z)$.

Wang tiles are unit squares with colored edges which may not be flipped or rotated. A *tileset* T is a finite set of Wang tiles. A *coloring of the plane* is a mapping $c : \mathbb{Z}^2 \to T$ assigning a Wang tile to each point of the plane. If all adjacent tiles of a coloring of the plane have matching edges, it is called a tiling.

In particular, the set of tilings of a Wang tileset is a SFT on the alphabet formed by the tiles. Conversely, any SFT is isomorphic to a Wang tileset. From a recursivity point of view, one can say that SFTs and Wang tilesets are equivalent. In this paper, we will be using both indiscriminately. In particular, we note X_T the SFT associated to a set of tiles T.

We say a SFT (tileset) is *origin constrained* when the letter (tile) at position $(0, 0)$ is forced, that is to say, we only look at the valid tilings having a given letter (tile) t at the origin.

More information on SFTs may be found in Lind and Marcus' book [LM95].

The notion of *Cantor-Bendixson derivative* is defined on set of configurations. This notion was introduced for tilings by Ballier/Durand/Jeandel [I7]. A configuration c is said to be *isolated* in a set of configurations C if there exists a pattern p such that c is the only configuration of C containing p. The Cantor-Bendixson derivative of C is noted $D(C)$ and consists of all configurations of C except the isolated ones. We define $C^{(\lambda)}$ inductively for any ordinal λ:

– $C^{(0)} = S$
– $C^{(\lambda+1)} = D\left(C^{(\lambda)}\right)$
– $C^{(\lambda)} = \bigcap_{\gamma<\lambda} C^{(\gamma)}$ when λ is limit.

The *Cantor-Bendixson rank* of C, noted $CB(C)$, is defined as the first ordinal λ such that $C^{(\lambda)} = C^{(\lambda+1)}$. An element x is of rank λ in C if λ is the least ordinal such that $x \notin C^{(\lambda)}$.

A configuration x is *periodic*, if there exists $n \in \mathbb{N}^\star$ such that $\sigma_{ne_i}(x) = x$, for any $i \in \{1, \ldots, d\}$, where the e_i's form the standard basis. A *vector of periodicity* of a configuration is a vector $v \in \mathbb{Z}^d \setminus \{(0, \ldots, 0)\}$ such that $\sigma_v(x) = x$. A configuration x is *quasiperiodic* (see Durand [Dur99] for instance) if for any pattern p appearing in x, there exists N such that this pattern appears in all N^d cubes in x. In particular, a periodic point is quasiperiodic. A configuration is *strictly quasiperiodic* if it is quasiperiodic and not periodic. A subshift is *minimal* if it is non-empty and contains no proper non-empty subshift. Equivalently, all its points have the same patterns. In this case, it contains only quasiperiodic points. It is known that every subshift contains a minimal subshift, see e.g. Durand [Dur99].

F.1.2 Computability background

A Π_1^0*class* $P \subseteq \{0,1\}^{\mathbb{N}}$ is a class of infinite sequences on $\{0, 1\}$ for which there exists a Turing machine that given $x \in \{0,1\}^{\mathbb{N}}$ as an oracle halts if and only if $x \notin P$. Equivalently, a class $S \subseteq \{0,1\}^{\mathbb{N}}$ is Π_1^0if there exists a recursive set L so that $x \in S$ if no prefix of x is in L. An element of a Π_1^0class is called a *member* of this class.

We say that two sets S, S' are *recursively homeomorphic* if there exists a bijective recursive function $f : S \to S'$. That is to say there are two Turing machines M (resp. M') such that given a member of S (resp. S') computes a member of S' (resp. S). Furthermore, for any $s \in S, s' \in S'$ such that s' is computed by M from s, M' computes s from s'.

The *Cantor-Bendixson rank* of S, is well defined similarly as for subshifts.

See Cenzer/Remmel [CR98] for Π_1^0sets and Kechris [Kec95] for Cantor-Bendixson rank and derivative.

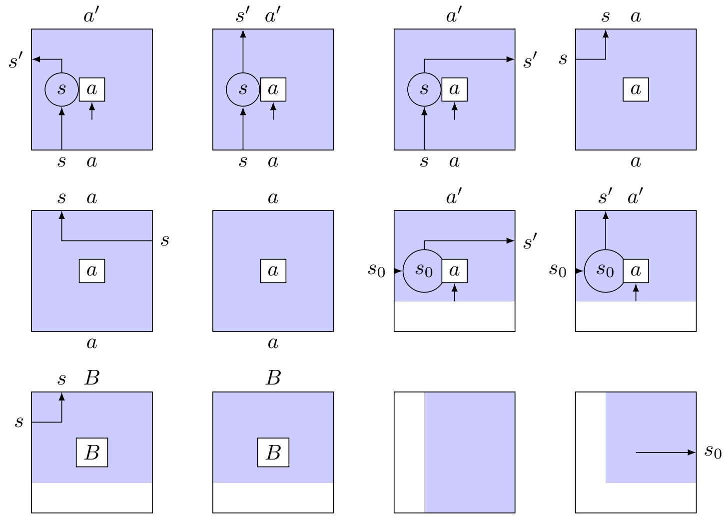


Figure F.1: A set of Wang tiles, encoding computation of a Turing machine: the states are in the circles and the tape is in the rectangles. A tiling containing the bottom right tile contains the space-time diagram of a run of the Turing machine.

For $x, y \in \{0,1\}^{\mathbb{N}}$ we say that x is *Turing-reducible* to y if y is computable by a Turing machine using x as an oracle and we write $y \leq_T x$. If $x \leq_T y$ and $y \leq_T x$, we say that x and y are *Turing-equivalent* and we write $x \equiv_T y$. The *Turing degree* of $x \in \{0,1\}^{\mathbb{N}}$ is its equivalence class under the relation $\equiv_T$.

F.1.3 Subshifts and Π_1^0classes

As it is clear from the definitions, SFTs in any dimension are Π_1^0classes. More generally, *effective* subshifts, see e.g. Cenzer/Dashti/King [CDK08]), that is subshifts defined by a computable (or equivalently, in this case, by a computably enumerable) set of forbidden patterns are Π_1^0classes. As such, they enjoy similar properties. In particular, there exists many "basis theorems", ie theorems that assert that any Π_1^0(non-empty) class has a member with some specific property.

As an example, every countable Π_1^0class has a computable member, see e.g. Cenzer/Remmel [CR11]. For subshifts, we can say a bit more: every countable subshift has a periodic (hence computable) member. Every Π_1^0class has a point of low degree, as prove in Jockusch/Soare [JS72c]. A proof of this from the point of view of subshifts (actually tilings) is given in Durand/Levin/Shen [DLS08].

F.2 Π_1^0sets and origin constrained tilings

A straighforward corollary of Hanf [Han74] is that Π_1^0classes and origin constrained SFTs are recursively isomorphic. This is stated explicitely in Simpson [Sim11b].

Theorem F.1

Given any Π^0_1class $P \subseteq \{0,1\}^{\mathbb{N}}$, there exists a SFT X and a letter $t \in \Sigma_X$ such that each origin constrained point corresponds to a member of P.

Proof : Let P be a Π^0_1class, and M the Turing machine that proves it, that is M given $x \in \{0,1\}^{\mathbb{N}}$ as an oracle halts if and only if $x \notin P$.

We use the classic encoding of Turing machines, see fig. F.1. We modify all tiles containing a symbol from the tape, to allow them to contain a second symbol. This symbol is copied vertically. All these second symbols represent the oracle.

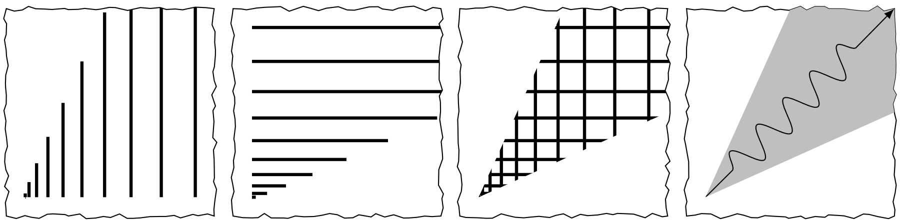
Then the SFT constrained by the tile starting the computation contains exactly the runs of the Turing machine with members of P on the oracle tape. ∎

Corollary F.2

Any Π^0_1subset P of $\{0,1\}^{\mathbb{N}}$ is recursively homeomorphic to an origin constrained SFT.

F.3 Producing a sparse grid

The main problem in the previous construction is that points which do not have the given letter at the origin can be very wild: they may correspond to configurations with no computation (no head of the Turing Machine) or computations starting from an arbitrary (not initial) configuration. A way to solve this problem is described in Myers' paper [Mye74] but is unsuitable for our purposes (It was however used by Simpson to obtain a weaker result on Medvedev degrees, see [Sim11b]).



(a) (b) (c) (d)

Figure F.2: The tiling in which the Turing machines will be encoded.

Our idea is as follows: we build a SFT which will contain, among others points, the *sparse grid* of Figure F.2c. The interest being that all others points will have at most one intersection of two black lines. This means that if we put computation cells of a given Turing machine in the intersection points, every point which is not of the form of Figure F.2c will contain at most one cell of the Turing machine, and thus will contain no computation.

To do this construction, we will first draw increasingly big and distant columns as in Figure F.2a and then superimpose the same construction for rows as in Figure F.2b, thus obtaining the grid of Figure F.2c.

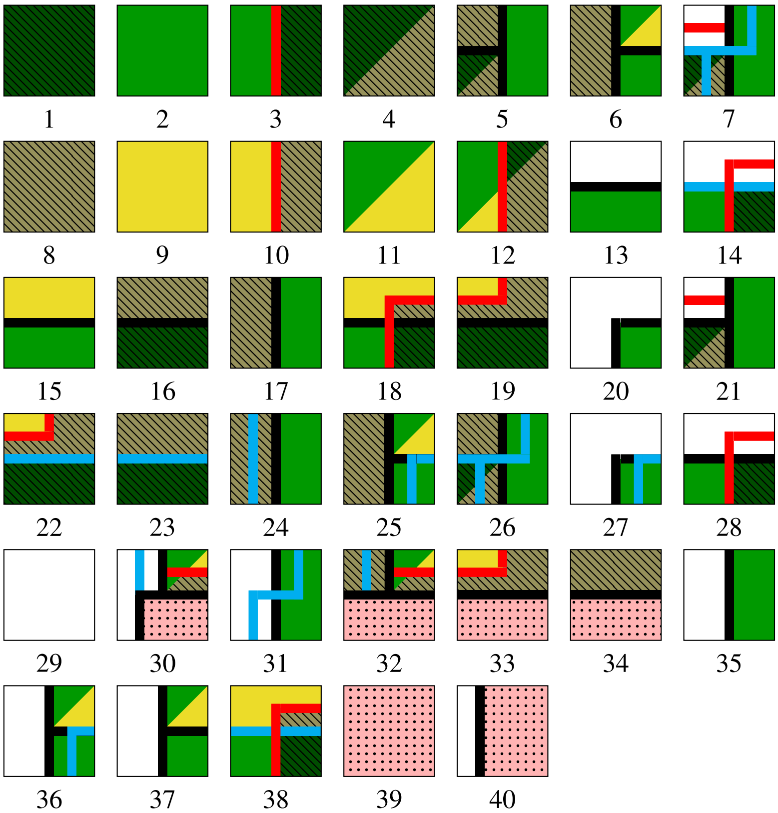


Figure F.3: Our set of Wang tiles T.

It is then fairly straightforward to see how we can encode a Turing machine inside a configuration having the skeleton of Figure F.2c by looking at it diagonally: time increases going to the north-east and the tape is written on the north-west/south-east diagonals [1].

Our set of tiles T of Figure F.3 gives the skeleton of Figure F.2a when forgetting everything but the black vertical borders. We will prove in this section that it is countable. We set here the vocabulary:

– a vertical line is formed of a vertical succession of tiles containing a vertical black line (tiles 5, 6, 17, 21, 24, 25, 26, 27, 31, 35, 36, 37).
– a horizontal line is formed of a horizontal succession of tiles containing a horizontal black line (tiles 13, 14, 15, 16, 22, 23, 38) or a bottom signal,
– the bottom signal is formed by a connected path of tiles among (30, 31, 27, 14, 7, 36, 38)
– the red signal is formed by a connected path of tiles containing a red line (tiles among 3 ,7, 10, 12, 14, 19, 22, 32, 33, 38).
– tile 30 is the corner tile
– tiles 30, 32, 33, 34 are the bottom tiles

Lemma F.3

The SFT X_T admits at most one point, up to translation, with two or more vertical lines. This point is drawn on Figure F.4.

Proof : The idea of the construction is to force that whenever there are two vertical lines, then the point is a shifted of the one in Figure F.4. Note also that whenever the corner tile appears in a point, it is necessarily a shifted version of the point on Figure F.4.

Suppose that we have a tiling in which two vertical lines appear. These two lines necessarily face each other horizontaly: it is impossible for them not to have a bottom, and their bottoms are at the same height. Suppose the horizontal distance between them is $k + 1$. There must then be horizontal lines between them forming squares, because of the diagonal. Inside these squares there must be a red signal: inside each square, this red signal is vertical, it is shifted to the right each time it crosses a horizontal line. This ensures that there are exactly k squares in this column. Furthermore, the bottom square has necessarily a bottom signal going through its top horizontal line. The bottom signal forces the square of the column before to be of size $k - 1$ and the square of the column after to be of size exactly $k + 1$. Thus, the corner tile appears in the point. ■

Lemma F.4

X_T is countable.

Proof : Lemma F.3 states that there is one point, up to shift, that has two or more vertical lines. This means that the other points have at most one such line.

– If a point has exactly one vertical line, then it can have at most two horizontal lines: one on the left of the vertical one and one on the right. A red signal can then appear on the left or the right of the vertical line arbitrary far from it. There is a countable number of such points.
– If a point has no vertical line, then it has at most one horizontal line. A red signal can then appear only once. There is a finite number of such points, up to shift.

1. Note that we will have to skip one diagonal out of two in our construction, in order for the tape to increase at the same rate as the time.

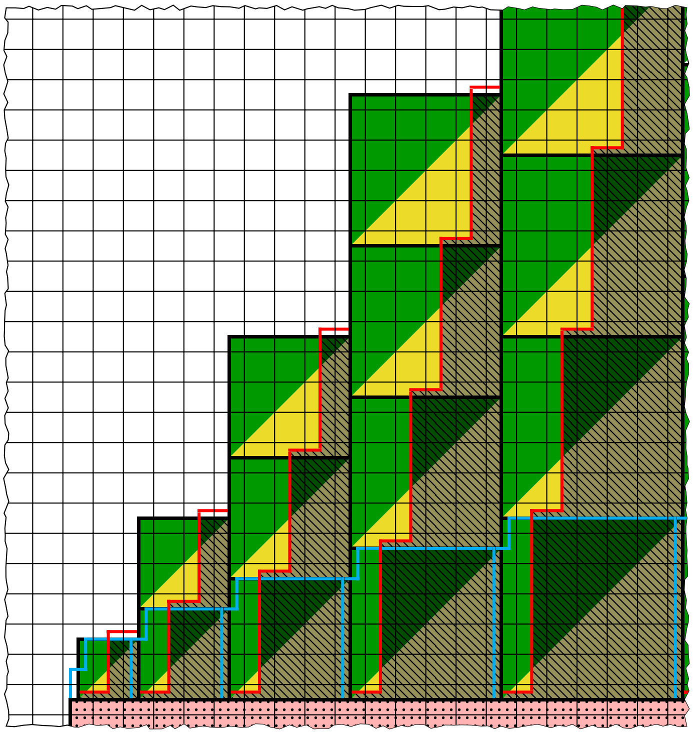


Figure F.4: Tiling α: the unique valid tiling of T in which there are 2 or more vertical lines.

There is a countable number of points that can be obtained with the tileset T. All types of obtainable points are shown in Figure F.5 and F.4. ■

By taking our tileset $T = \{1, \ldots, 40\}$ and mirroring all the tiles along the south-west/north-east diagonal, we obtain a tileset $T' = \{1', \ldots, 40'\}$ with the exact same properties, except it enforces the squeleton of Figure F.2b. Remember that whenever the corner tile appeared in a point, then necessarily this point was a shifted of α. Analogously, the corner tile of T' appearing in a point means the this point is a shifted of α'. We hence construct a third tileset $\tau = (T \setminus \{30\} \times T' \setminus \{30'\}) \cup \{(30, 30')\}$ which is the superimposition of T and T' with the restriction that tiles 30 and $30'$ are necessarily superimposed to each other. The corner tile $(30, 30')$ of τ has the property that whenever it appears, the tiling is the superimposition of the skeletons of Figures F.2a and F.2b with the corner tiles at the same place: there is only one such tiling, we call it β.

The skeleton of Figure F.2c is obtained from β if we forget about the parts of the lines of the T layer (resp. T') that are superimposed to white tiles, 29' (resp. 29), of T' (resp. T).

As a consequence of Lemma F.4, X_τ is also countable. And as a consequence of Lemma F.3, the only points in x_τ in which computation can be embedded are the shifts of β. The shape of β is the one of Figure F.2c, the coordinates of the points of the grid are the following (supposing tile $(30, 30')$ is at the center of the grid):

$$\{(f(n), f(m)) \mid f(m)/4 \leq f(n) \leq 4f(m)\}$$

$$\{(f(n), f(m)) \mid m/2 \leq n \leq 2m\}$$

where $f(n) = (n+1)(n+2)/2 - 1$.

Lemma F.5

The Cantor-Bendixson rank of X_τ is 12.

Proof : The Cantor-Bendixson rank of $X_T \setminus \{\alpha\}$ is 6, see Figure F.5, thus the rank of $X_T \setminus \{\alpha\} \times X_{T'} \setminus \{\alpha'\}$ is 11. Adding the configurations corresponding to the superimposition of α and α', X_τ has rank 12. ■

F.4 Π^0_1classes with recursive members and SFTs

The SFT constructed before will allow us to prove a series of theorems on Π^0_1classes with recursive points. The foundation of these is Theorem F.6 which establishes a recursive homeomorphism between SFTs and Π^0_1classes, up to a recursive subset of the SFT. This recursive homeomorphism is the best we can hope for, as will be shown in section F.5. Then from this "partial" homeomorphism, we will be able to deduce results on the set of Turing degrees of SFTs and Π^0_1classes.

Theorem F.6

For any Π^0_1class S of $\{0,1\}^{\mathbb{N}}$ there exists a tileset τ_S such that $S \times \mathbb{Z}^2$ is recursively homeomorphic to $X_{\tau_S} \setminus O$ where O is a computable set of configurations.

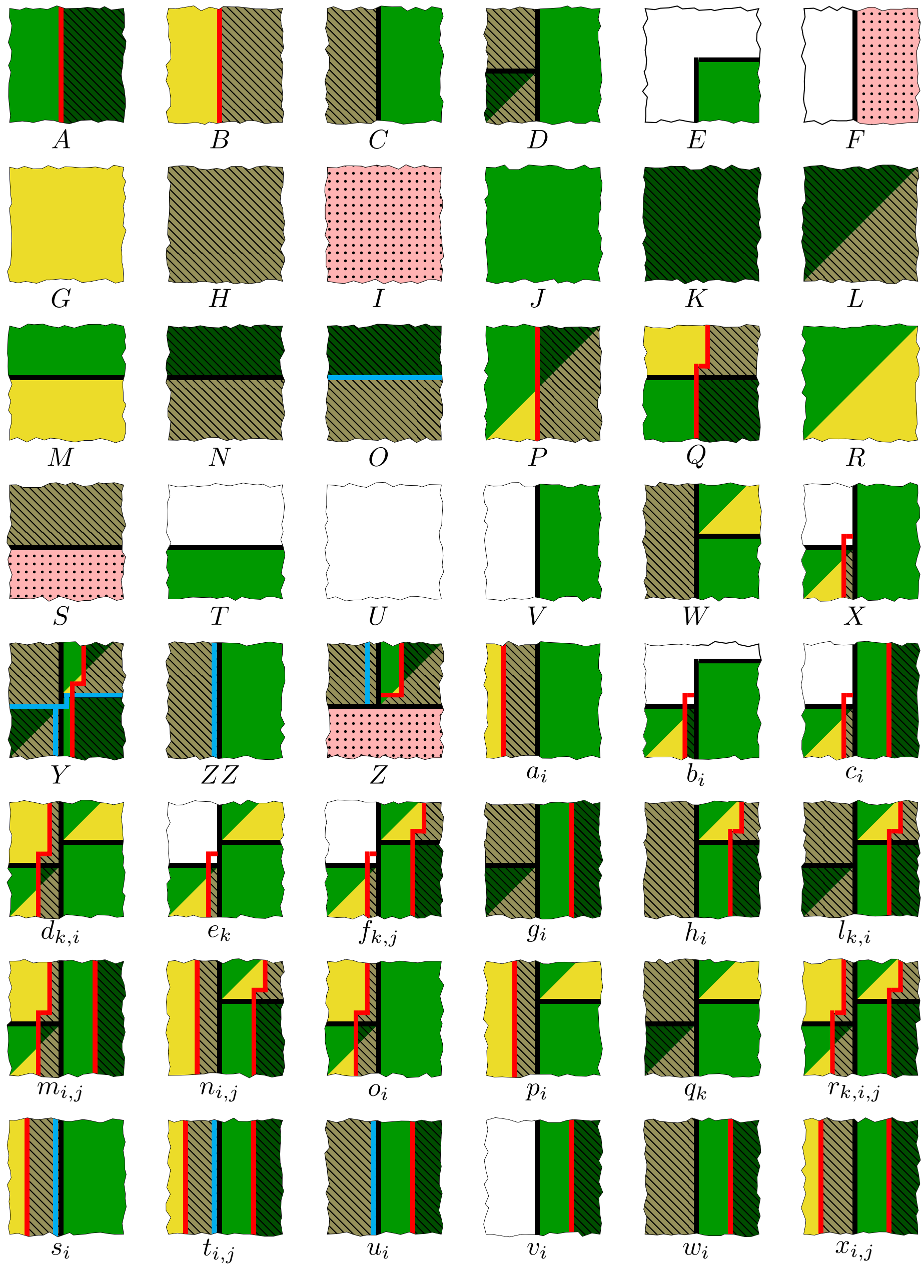


Figure F.5: The other configurations: the $A - ZZ$ configurations are unique (up to shift), and the configurations with subscripts $i, j \in \mathbb{N}, k \in \mathbb{Z}^2$ represent the fact that distances between some of the lines can vary. Note that configuration ZZ cannot have a red signal on its left, because it would force another vertical line.

Proof : This proof uses the construction of section F.3. Let M be a Turing machine such that M halts with x as an oracle iff $x \notin S$. Take the tileset τ of section F.3 and encode, as explained earlier, in configuration β the Turing machine M having as an oracle x on an unmodifiable second tape. This gives us τ_M, O is the set all points except the β ones. To each $(x, z) \in S \times \mathbb{Z}^2$ we associate the β tiling having a corner at position z and having x on its oracle tape. O is computable, because it contains a countable number (Lemma F.4) of computable points (none of these points can contain more than one step of computation). ■

Corollary F.7

For any Π_1^0class S of $\{0,1\}^{\mathbb{N}}$ with a recursive member, there exists a SFT X with the same set of Turing degrees.

Corollary F.8

For any countable Π_1^0class S of $\{0,1\}^{\mathbb{N}}$, there exists a SFT X with the same set of Turing degrees.

Proof : We know, from Cenzer/Remmel [CR98], that countable Π_1^0sets have **0** (computable elements) in their set of Turing degrees, thus the SFT X_{τ_M} described in the proof of Theorem F.6 has exactly the same set of Turing degrees as S. ■

Theorem F.9

For any countable Π_1^0class S of $\{0,1\}^{\mathbb{N}}$ there exists a SFT X with the same set of Turing degrees such that $CB(X) = CB(S) + 11$.

Proof : Lemma F.5 states that X_τ is of Cantor-Bendixson rank 12, 11 without β. In the tileset τ_M of the previous proof, the Cantor-Bendixson rank of the contents of the tape is exactly $CB(S)$, hence $CB(X_{\tau_S}) = CB(S) + 11$. ■

From Ballier/Durand/Jeandel [I7] we know that for any subshift X, if $CB(X) \leq 2$, then X has only recursive points. Thus an optimal construction would have to augment the Cantor-Bendixson rank by at least 2.

Corollary F.10

For any countable Π_1^0class S of $\{0,1\}^{\mathbb{N}}$ there exists a sofic subshift X with the same set of Turing degrees such that $CB(X) = CB(S) + 2$.

Proof : Take a projection that just keeps the symbols of the Turing machine tape τ_M of the proof of Theorem F.6 and maps everything else to a blank symbol. Recall the Turing machine tape cells are the intersections of the vertical lines and horizontal lines. This projection leads to 3 possible configurations :

- a completely blank configuration,
- a completely blank configuration with only one symbol somewhere,
- a configuration with a white background and points corresponding to the intersections in the sparse grid of Figure F.2c.

■

Note that a similar theorem in dimension one for effective rather than sofic subshifts was conjectured in Cenzer/Dashti/Toska/Wyman [CDTW10] and later proved by the authors (personal communication).

F.5 Π^0_1classes without recursive members and subshifts

In this section we prove that two-dimensional SFTs containing only non-recursive points have the property that they always have points with different but comparable degrees, this is corollary F.20. But we first prove this result for one-dimensional subshifts, not necessarily of finite type, in Theorem F.13, the proof for two-dimensional SFTs needing only a bit more work.

One interest of these proofs, lies in the following theorem, proved by Jockusch and Soare:

Theorem F.11 (*Jockusch, Soare*)

> There exists Π^0_1classes containing no recursive member, such that any two different members are Turing-incomparable.

The proof of this result can be found in Cenzer and Remmel's upcoming book [CR98] or in the original articles by Jockusch and Soare [JS72c, JS72a].

This means that one cannot expect a full recursive homeomorphism, i.e. without removal of the recursive points. Furthermore, this shows that in general, when a Π^0_1class P has no computable member, it is not true that one can find a SFT with the same set of Turing degrees.

The main idea of the proof is that any subshift contains a minimal subshift. If the subshift has no recursive points (actually, no periodic points), this minimal subshift contains only strictly quasiperiodic points. We will then use some combinatorial properties of this minimal subshift to obtain our results.

F.5.1 One-dimensional subshifts

We start with a technical lemma that will allow us to prove the theorem:

Lemma F.12

> Let x be a strictly quasiperiodic point of a minimal one-dimensional subshift A and $\prec$ be an order on Σ_A. For any word w extensible to x, there exists two words w_0 and w_1 such that:
> - w appears exactly twice in w_0 and w_1 respectively,
> - let a and b (resp. c and d) be the first differing letters in the blocks directly following the first and second occurence of w in w_0 (resp. w_1), then $a \prec b$ (resp. $d \prec c$).

Proof : By quasiperiodicity of x, w appears infinitely many times in x. By non periodicity, any two occurences of w must be followed by eventually distinct words. Let y be the largest word so that whenever w appears in x, then wy appears. Note that w appears only once in wy, otherwise the x would be periodic.

By definition of y, the letters after each occurence of wy cannot be all the same. So there exists two consecutive occurences of wy with differing next letters a, b with, e.g., $a \prec b$ (the other case being similar). w_0 is then defined as the smallest word containg both occurences of wy and these letters a, b.

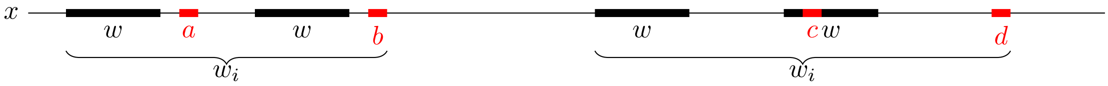


Figure F.6: Two nearest w blocks, the first differing letter a, b and c, d in their following blocks, and how they form the w_i. Note that the first differing letter might in some cases be inside the second occurence of w, as illustrated on the right with c, d.

Now x is quasiperiodic, hence some occurence of wyb must also appear before some occurence of wya, so we can find between these two positions two occurences of wy with differing next letters c, d with $d \prec c$. We can then define w_1 similarly.

See Figure F.6 for an illustration of the construction of w_0 and w_1. ■

Theorem F.13

Let A be a minimal subshift containing only strictly quasiperiodic points and x a point of A. Then for any Turing degree d such that $\deg_T x \leq d$, there exists a point $y \in A$ with Turing degree d.

Proof : To prove the theorem, we will give two computable functions $f : A \times \{0,1\}^{\mathbb{N}} \to A$ and $g : A \to \{0,1\}^{\mathbb{N}}$ such that for any $x \in A$ and $s \in \{0,1\}^{\mathbb{N}}$ we have $g(f(x,s)) = s$. This means in terms of Turing degrees:

$$\deg_T s \leq \deg_T f(x,s) \leq \sup_T(\deg_T x, \deg_T s)$$

That is to say, we give two algorithms, one (f) that given a point x of A and a sequence s of $\{0,1\}^{\mathbb{N}}$ reversibly computes a point of A that embeds s, the second (g) retrieves s from the computed point.

Let us now give f. Let $\prec$ be an order on Σ_A. Given a point $x \in A$ and a sequence $s \in \{0,1\}^{\mathbb{N}}$, f recursively constructs another point of A: it starts with a block $C_{-1} = x_0$ and recursively constructs bigger and bigger blocks C_i such that C_{i+1} has C_i in its center. Furthermore these blocks are each centered in 0. So that the sequence $C_{-1} \to C_0 \to C_1 \to \cdots \to C_i \to \ldots$ converges to a point c of A having all C_i's in its center. It is sufficient to show then how C_{i+1} is constructed from C_i.

f works as follows: It searches for two consecutive occurences of C_i in x, where the two first differing letters satisfy $a \prec b$ if $s_{i+1} = 0$ and $b \prec a$ if $s_{i+1} = 1$. We know that f will eventually succeed in finding these occurences due to Lemma F.12.

Now we define C_{i+1} as the word in x where we find these two occurences, correctly cut so that the first occurence of C_i is at its center, and its last letter is the differing letter of the second occurence. See Figure F.7.

We thus have f, which is clearly computable. We give now g.

Given C_i and c, one can compute s_{i+1} easily: we just have to look for the second occurence of C_i in c, the first one being in its center. We then check whether the first differing letters between the blocks following each occurence are such that $e \prec f$ or $f \prec e$. This also gives us C_{i+1}.

This means that from c, one can recover s. We know $C_{-1} = c_0$ and from this information, we can get the rest: from c and C_i, one computes easily C_{i+1} and s_i. We have constructed our function g.

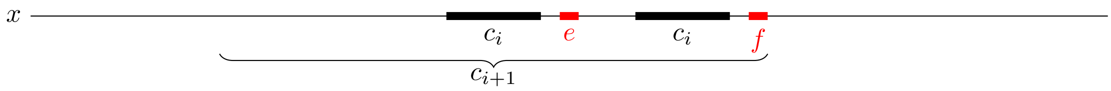


Figure F.7: How we construct c_{i+1} from c_i. When $s_{i+1} = 0$, we have $e \prec f$ and $f \prec e$ otherwise. The words of Lemma F.12 are completed on the left with the block preceeding them in x.

So now if we take a sequence s such that $\deg_T s > \deg_T x$, we can take $y = c = f(x, s)$. It has the same Turing degree as s since $\deg_T s = \sup_T(\deg_T x, \deg_T s)$. ■

Corollary F.14

Every non-empty one-dimensional subshift S containing only non recursive points has points with different but comparable degrees.

Proof : Take any minimal subshift of S. It must contain only strictly quasiperiodic points, so the previous theorem apply. ■

For effective subshifts, we can do better

Lemma F.15

Every nonempty effective subshift S contains a minimal subshift $\tilde{S}$ whose set of valid patterns is of Turing degree less than $0'$.

$0'$ is the degree of the Halting problem.

Proof : Let $\mathcal{F}$ be the computable set of forbidden patterns defining S. Let w_n be a (computable) enumeration of all words. Define $\mathcal{F}_n$ as follows: $\mathcal{F}_{-1} = \emptyset$. Then if $\mathcal{F}_n \cup \mathcal{F} \cup \{w_{n+1}\}$ defines a non-empty subshift, then $\mathcal{F}_{n+1} = \mathcal{F} \cup \{w_{n+1}\}$ else $\mathcal{F}_{n+1} = \mathcal{F}_n$.

Now take $\tilde{\mathcal{F}} = \cup_n \mathcal{F}_n$. It is clear from the construction that $\tilde{\mathcal{F}}$ is computable given the Halting problem. Moreover $\tilde{\mathcal{F}}$ defines a non-empty, minimal subshift $\tilde{S}$. More exactly the complement of $\tilde{\mathcal{F}}$ is exactly the set of patterns appearing in $\tilde{S}$. ■

This lemma cannot be improved: an effective subshift is built in Ballier/Jeandel [I3] for which the language of every minimal subshift is at least of Turing degree $0'$.

Now it is clear that any minimal subshift $\tilde{S}$ has a point computable in its set of valid patterns, so that:

Corollary F.16

Every non-empty effective subshift with no recursive point contains configurations of every Turing degree above $0'$.

We do not know if this can be improved. While it is true that all minimal subshifts in [I3] have a language of Turing degree at least $0'$, this does not mean that their configurations have all Turing degree at least $0'$. In the construction of [I3], there indeed exists recursive minimal points. The construction of Myers [Mye74] has nonrecursive points, but points of low degree.

F.5.2 Two-dimensional SFTs

We now prove an analoguous theorem for two dimensional SFTs. We cannot use the previous result directly as it is not true that any strictly quasiperiodic configuration always contain a strictly quasiperiodic (horizontal) line. Indeed, there exists strictly quasiperiodic configurations, even in SFTs with no periodic configurations, where some line in the configuration is not quasiperiodic (this is the case of the "cross" in Robinson's construction [Rob71]) or for which every line is periodic of different period (such configurations happen in particular in the Kari-Culik construction [II96, Kar96]).

We will first try to prove a result similar to Lemma F.12, for which we will need an intermediate definition and lemma.

Definition F.1 (*line*)

A *line* or n-*line* of a two-dimensional configuration $x \in \Sigma^{\mathbb{Z}^2}$ is a function $l : \mathbb{Z}\times H \to \Sigma$, with $H = h + [\![0; n-1]\!]$, $h \in \mathbb{Z}$, such that

$$x_{|\mathbb{Z}\times H} = l.$$

Where n is the width of the line and h the vertical placement.

One can also define a line in a block by simply taking the intersection of both domains. The notion of quasiperiodicity for lines is exactly the same as the one for one dimensional subshifts. We need this notion for the following lemma, that will help us prove the two-dimensional version of Lemma F.12. We also think that this lemma might be of interest in itself.

Lemma F.17

Let A be a minimal subshift. There exists a point $x \in A$ such that all its lines are quasiperiodic.

Proof : Let $\{(a_i, b_i)\}_{i\in\mathbb{N}}$ be an enumeration of $\mathbb{Z} \times \mathbb{N}$ and $H_i = a_i + [\![0; b_i]\!]$.

If x is a configuration, denote by $p_i(x) : \mathbb{Z} \times H_i \mapsto \Sigma$ the restriction of x to $\mathbb{Z} \times H_i$. We will often view p_i as a map from A to $(\Sigma^{H_i})^{\mathbb{Z}}$. A *horizontal* subshift is a subset of $\Sigma^{\mathbb{Z}^2}$ which is closed and invariant by a horizontal shift.

We will build by induction a non-empty horizontal subshift A_i of A with the property that every configuration x of A_i has the property that every line of support H_j, for any $j < i$, is quasiperiodic. More precisely, $p_j(A_i)$ will be a minimal subshift.

Define $A_{-1} = A$. If A_i is defined, consider $p_{i+1}(A_i)$. This is a non-empty subshift, so it contains a minimal subshift X. Now we define the horizontal subshift $A_{i+1} = p_{i+1}^{-1}(X) \cap A_i$. By construction $p_{i+1}(A_{i+1})$ is minimal. Furthermore, for any $j < i$, $p_j(A_{i+1})$ is a non-empty subshift, and it is included in $p_j(A_j)$, which is minimal, hence it is minimal.

To end the proof, remark that by compactness $\cap_i A_i$ is non-empty, as every finite intersection is non-empty.

■

Lemma F.18

Let A be a two-dimensional minimal subshift where all points (equivalently, some point) have no horizontal period.

Let x be a point of A and $\prec$ be an order on Σ_A. For each $n \in \mathbb{N}$, for any n-block w extensible to x, there exists two blocks w_0 and w_1 extensible to x such that:

- w appears exactly twice in both w_0 and w_1, on the same n-line of vertical placement 0.
- the first differring letters e and f in the blocks containing and starting with each occurence of w are such that $e \prec f$ in w_0 and $f \prec e$ in w_1.

Here the word “first” refers to an adequate enumeration of $\mathbb{N} \times \mathbb{Z}$.

Proof : As the result is about patterns rather than configurations, we can suppose w.l.o.g by Lemma F.17 that all lines of x are quasiperiodic.

Since w appears in x, it appears a second time on the same n-line in x. Since x is not horizontally periodic, both occurences are in the center of different blocks. (The place where they differ may be on a different line, though, if this particular n-line is periodic)

Now we use the same argument as lemma F.12 on the m-line containing both occurences of w and the first place they differ. (Note that we cannot use directly the lemma as this m-line might itself be periodic, but the proof still works in this case) ■

Theorem F.19

Let A be a two-dimensional minimal subshift where all points (equivalently, some point) have no horizontal period and x a point of A. Then for any Turing degree d such that $\deg_T x \leq d$, there exists a point $y \in A$ with Turing degree d.

Proof : The proof is identical as the one of Theorem F.13, Lemma F.18 being the two-dimensional counterpart of Lemma F.12. One can see in Figure F.8 how the C_i's are contructed in this case. ■

Corollary F.20

Every two-dimensional non-empty subshift X containing only non-recursive points has points with different but comparable degrees.

Proof : X contains a minimal subshift A, which cannot be periodic. If A contains a point with a horizontal period, then all points of A have a horizontal period, and the result follows from Theorem F.13. Otherwise, it follows from the previous theorem. ■

Lemma F.15 is still valid in any dimensions so that we have:

Corollary F.21

Every two-dimensional non-empty effective subshift (in particular any non-empty SFT) with no recursive points contains points of any Turing degree above $0'$.

We conjecture that a stronger statement is true: The set of Turing degrees of any subshift with no recursive points is upward closed. To prove this, it is sufficient to prove that for any subshift S and any configuration x of S (which is not minimal), there exists a minimal configuration in S

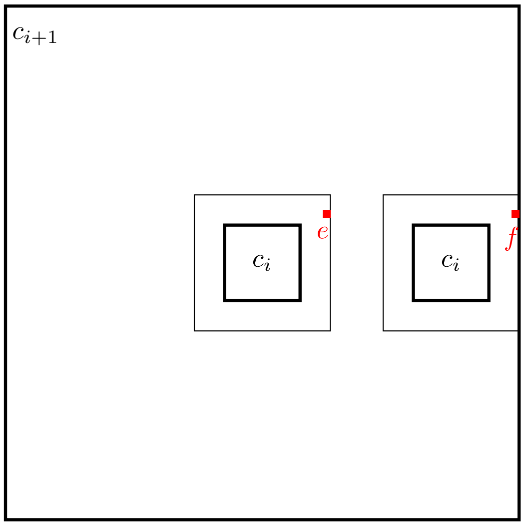


Figure F.8: How c_{i+1} is constructed inductively from c_i. c_i is in the center of c_{i+1}. The letters e and f are the first differing letters in the blocks containing the c_i's. Whether $e \prec f$ of $f \prec e$ depends on what symbol we want to embed, 0 or 1.

of Turing degree less than the degree of x. We however have no idea how to prove this, and no counterexample comes to mind.

Annexe G

TILINGS AND MODEL THEORY

L'article [I8], écrit en collaboration avec Alexis Ballier, esquisse les liens entre pavages et théorie des modèles. Il fait ainsi echo au chapitre 4.

Abstract

In this paper we emphasize the links between model theory and tilings. More precisely, after giving the definitions of what tilings are, we give a natural way to have an interpretation of the tiling rules in first order logics. This opens the way to map some model theoretical properties onto some properties of sets of tilings, or tilings themselves.

G.1 Introduction

Tilings are a basic and intuitive way to express geometrical constraints; they happened to be of broad interest in computer science since Berger proved the undecidability of the domino problem [Ber64] by showing that they can embed, despite being static objects, some kind of computation. This also was the first step in the links between logics and tilings as they helped to prove the undecidability of some classes for formulae [BGG01, Wan60, Wan61, Wan63]. Some more links have then been discovered by Makowsky that used previous constructions of aperiodic tilesets to show the existence of a complete, finetely axiomatizable and superstable theory [Mak74]. Some recent results by Oger generalize this approach to more abstract definitions of tilings and proves some nice equivalences between model theory and this generalized definition [Oge04].

In this paper we will give details of constructions used to translate tilings and tileset properties into model theoretic ones. Section G.2 will be devoted to the proper definitions of tilings and tilesets; We will then translate these definitions into first order formulae in Section G.3. Finally in Section G.4 we shall present the equivalence results that can be obtained by this translation.

Most of these results are already present in [Mak74, Poi80]. However we hope that this paper will offer a new look at these results.

The major part of this paper is devoted to tilings of the plane $\mathbb{Z}^2$. However, we may define similar theories for tilings of other spaces such as $\mathbb{Z}^3$ or any Cayley graph. The article [Oge04] in particular deals with tilings of $\mathbb{R}^n$ by polytopes.

G.2 Tilings

Several definitions of discrete tilings can be found in the litterature, but are equivalent for many purposes [Cer02]. We will focus here on the definition by forbidden patterns.

First we have to define the space we are going to tile: we want to assign a state taken in a finite set Q to each cell of the discrete plane $\mathbb{Z}^2$. Q may be seen as a set of colors, or a set of states. Therefore, we define the set of configurations as the functions from $\mathbb{Z}^2$ to Q:

Definition G.1

The set of configurations is $Q^{\mathbb{Z}^2}$.

The patterns are nothing but a configuration restricted to a finite domain; that is, considering a finite subset D of $\mathbb{Z}^2$, a pattern is a function from D to Q.

Definition G.2

A pattern defined on a finite subset D of $\mathbb{Z}^2$ is an element of Q^D.

Informally a tileset represents geometric constraints imposed to the configurations, that is how the states in the cells of the plane are constrained by their neighborhood and how they constrain it. Formally we will define a valid tiling as a configuration that contain no forbidden pattern:

Definition G.3

A tileset is defined by a finite set of forbidden patterns $\mathcal{F}_\tau$.

A configuration c contains a pattern P defined on D (or equivalently P appears in c) if there exists $x \in \mathbb{Z}^2$ such that:

$$\forall y \in D, c(x+y) = P(y)$$

A configuration is said to be a valid tiling by τ if it contains no pattern in $\mathcal{F}_\tau$.

The so-called domino problem is to know given a tileset whether it generates a valid tiling. The problem has been proven undecidable by Berger in [Ber64].

We will now define a preorder $\preceq$ on configurations that focuses on patterns contained in them. This preorder has been defined in [Dur99, I7], however references to the concept can be found as early as [Poi80]:

Definition G.4 (*The pre-order* $\preceq$)

Let x, y be two configurations, we say that $x \preceq y$ if any pattern that appears in x also appears in y.

This induces the notion of local isomorphism between two configurations:

Definition G.5 (*Local isomorphism*)

Two configurations x and y are said to be locally isomorphic if $x \preceq y$ and $y \preceq x$. That is x and y contain the same patterns. We denote it by $x \approx y$.

Two configurations that are equal up to shift are locally isomorphic but the converse is not always true: there exists configurations that are locally isomorphic but one is not a shifted form of the other.

G.3 From tilesets to model theory

In this section we translate the definitions given in Section G.2 into first order formulae on some given language. This translation maps some properties of tilings onto some other properties of first order logics.

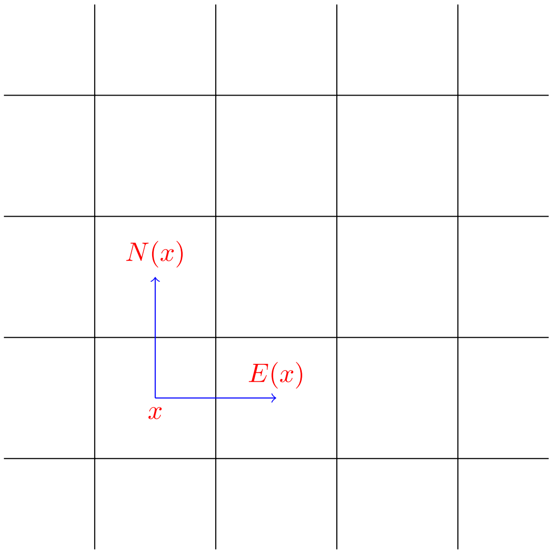


Figure G.1: The model we would like to obtain

Such a correspondence between tilesets and first order logic has already been defined [Poi80, Mak74] to show an example of finitely axiomatizable and superstable theory. A similar approach (see G.3.4) has been used to prove the undecidability of certain classes of formulae [Wan63, Wan60, Wan61, BGG01].

G.3.1 Axiomatizing the plane

The ideal model we would like to obtain is the plane $\mathbb{Z}^2$ like depicted in Figure G.1. The natural way to define cells on the plane $\mathbb{Z}^2$ is to consider them as variables and the adjacency relations between them as functions that allow us to move north, south, east or west from a given cell:

Definition G.6

We consider the language with the unary functions for movements on the plane: $\mathcal{L}_0$ is a set of unary functions : $\mathcal{L}_0 = \{N, S, E, W\}$.

And the corresponding axioms of the plane $\mathbb{Z}^2$:

– $\forall x, N(S(x)) = S(N(x)) = E(W(x)) = W(E(x)) = x$
– $\forall x, N(E(x)) = E(N(x))$

These formulae tends to axiomatize $\mathbb{Z}^2$ as a Cayley graph with two generators, the first formula axiomatizing the invertibility of the movements and the second the commutativity. However, these axioms are not sufficient, as we will see in the following sections.

Non standard models

With the axioms of the plane from the previous sections it is still possible to obtain some weird models. First, they also axiomatize some finite models like $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$, or some cylindric models like $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}$ (like e.g., in Figure G.2).

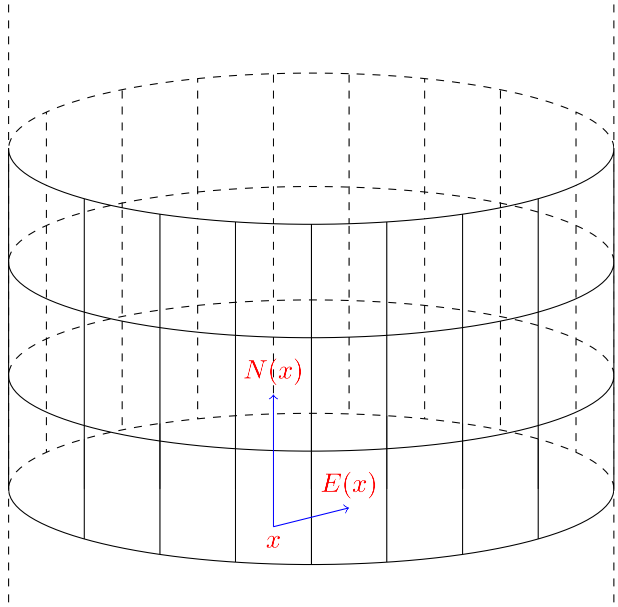


Figure G.2: A cylindric model

This problem can be dealt with by adding more axioms : For any i and j we may add the axiom $\forall x, E^i N^j(x) \neq x$. The main problem is then that the number of axioms is not finite, so that (we can prove that) the theory we obtain is not finitely axiomatisable anymore. However in most cases, the presence of these models is not a problem as we can "unfold" them into a plane (see e.g proof of lemma G.2).

Connectedness

The main problem however, which cannot be avoided, is that there is no way to ensure that all models of our theory are connected : A model is said to be connected if any two points can be connected using the N, S, E, W functions. An example of a disconnected model of our theory is depicted on Figure G.3. These disconnected models cannot be avoided. This is e.g., a consequence of the Löwenheim-Skolem theorem (There exist models of our theory of arbitrary infinite cardinals, these models cannot be connected if they are not countable) or more simply can be proven by a simple compactness argument: Consider a theory T that axiomatises the plane $\mathbb{Z}^2$. Add two constants c, d and the formulae ϕ_n that express that the points c and d are at distance at least n. Consider the theory $T' = T \cup \{\phi_n \mid n \in \mathbb{N}\}$. Any arbitary finite part of T' admits $\mathbb{Z}^2$ as a model (choose two points c and d arbitrary far) so that T' itself has a model by compactness. Such a model cannot be connected.

This proof also hints to a way to partially solve the problem. Consider formulae $\phi_n(x,y)$ that express that the points x and y are at distance at most n. Now consider the collection $p(x,y) = \{\phi_n(x,y), n \in \mathbb{N}\}$. $p(x,y)$ is a type, that is we can find for every finite part $q(x,y)$ of $p(x,y)$ some points c and d in any model so that $q(c,d)$ is true. Now we are interested in those models where p is not satisfied, that is in models where there do not exists c and d such that $p(c,d)$ is true. We say that such a model *omits* p

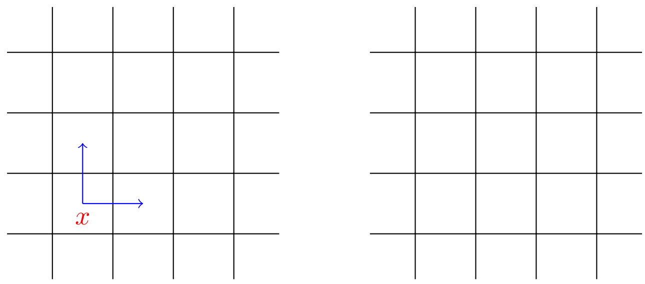


Figure G.3: An example of disconnected model

A part of model theory is devoted to the study of models omitting types. As an example, the omitting type theorem states that given a theory T any reasonable type can be omitted. However, most of the classical results in model theory will not work in this context, as e.g. the compactness theorem.

G.3.2 Encoding configurations

Now that we have some kind of axioms for the plane $\mathbb{Z}^2$, we may define what a configuration is. We defined a configuration as an application from $\mathbb{Z}^2$ to a finite set of states Q. We can code the states of the cells in our theory by unary predicates : we take one predicate Q_i for each state. The only thing we need to ensure is that each cell has exactly one state:

Definition G.7

New language:

$$\mathcal{L}_Q = \mathcal{L}_0 \cup \{Q_1, \ldots, Q_n\}$$

New axioms:

$$A : \forall x, \bigvee_i Q_i(x)$$

$$B : \forall x, \bigwedge_{j \neq i} (Q_i(x) \Rightarrow \neg Q_j(x))$$

We can also reduce the number of predicates by coding the states in binary form: for example, with 4 predicates, we can code up to 16 states.

G.3.3 The theory of a tileset

Following our definitions of tilesets in Section G.2, all what we need to do in order to encode them in first order logic is to write formulae that express "some specific pattern never appears". It can be done in the following way : Given a pattern P of domain D, any point p in D can be represented by a function that is a composition of the functions N, S, E, W. We can then write formulae that express that P appears at a point x:

Definition G.8

A formula to express that a pattern P defined on D appears at point x
$\varphi_P(x) := \bigwedge_{(i,j)\in D} P(i,j)(E^i(N^j(x)))$

As an example, the formula $\varphi = Q_1(E(x)) \wedge Q_2(x)$ expresses that x is of color 2 and its east neighbour is of color 1.

Then the formula $\forall x, \neg\varphi_P(x)$ axiomatizes that P never appears.

Definition G.9

> The theory T_τ of a tileset τ is the theory over the langage $\mathcal{L}_Q$ that contains all previous formulae and the formula $\forall x, \neg\varphi_P(x)$ for each forbidden pattern P. If the set of forbidden pattern $\mathcal{F}_\tau$ is finite, this theory is finitely axiomatisable.

G.3.4 Other languages

Before proceeding to the results, we give in this section various other languages in which to express tilings.

Another way to represent tilings is with a single unary function s (that intuitively denotes the successor of an integer) and with binary predicates P_i. $P_i(x, y)$ means that the state in the cell (x, y) is i. A structure is then over $\mathbb{Z}$ rather than $\mathbb{Z}^2$.

It is easy to represent forbidden patterns in this language. As an example, the formula $\phi = \forall x, y, \neg(P_1(x, y) \wedge P_2(s(x), y))$ means that there cannot be a cell in state 1 at the left of a cell in state 2.

Now suppose that the set of forbidden patterns has some particular form, that is constraints only concern adjacent cells. We now have a set of horizontal constraints H ($(i, j) \in H$ if a cell in state i cannot be at the left of a cell in state j) and vertical constraints V.

Now, the constraints can be written in the following way:

$$\phi = \forall x \forall y \bigwedge_{(i,j)\in H} (P_i(x, y) \Rightarrow \neg P_j(s(x), y)) \wedge \bigwedge_{(i,j)\in V} (P_i(x, y) \Rightarrow \neg P_j(x, s(y))$$

This can be rewritten (by a slight change of variables in the second part of the formula):

$$\forall x \forall y \bigwedge_{(i,j)\in H} (P_i(x, y) \Rightarrow \neg P_j(s(x), y)) \wedge \bigwedge_{(i,j)\in V} (P_i(y, x) \Rightarrow \neg P_j(y, s(x))$$

Now by a straightforward application of the skolemization process, we can replace the function s by a quantifier:

$$\forall x \exists x' \forall y \bigwedge_{(i,j)\in H} (P_i(x, y) \Rightarrow \neg P_j(x', y)) \wedge \bigwedge_{(i,j)\in V} (P_i(y, x) \Rightarrow \neg P_j(y, x')$$

We then obtain a new formula ϕ such that ϕ has a model if and only if there exists a tiling of the plane by the tileset. The proof proceeds as in lemma G.2 below. Note that the unfolding gives us only a tiling of a quarter of the plane. But it is known that a tileset can tile the entire plane if and only if it can tile a quarter of the plane.

The new formula ϕ is a formula with only three quantifiers $\forall\exists\forall$ and which contains only binary predicates. Thus we actually have proven that the class of formulae $[\forall\exists\forall, (0, \omega)]$ is undecidable. This is the core of the works by Wang, Kahr, Büchi about decidability of class of formulae. We then can deduce by an intricate transformation that the Kahr class $[\forall\exists\forall, (\omega, 1)]$ (one binary predicate, a finite number of unary predicates) is also undecidable.

See [Wan60, Wan61, Wan63] for more details. The encoding also has another property: The formula ϕ has a finite model if and only if there exists a periodic tiling of the plane by the tileset. This actually proves that the class $[\forall\exists\forall, (0, \omega)]$ is a *conservative reduction class*. See [BGG01] for more details.

G.4 Translating tilesets and tilings properties in model theoretical ones

We now show the links between those two different approaches.

Lemma G.1

A configuration can be seen as a structure over $\mathcal{L}_Q$. A tiling by τ can be seen as a model of T_τ.

This lemma is a consequence of the definitions we have taken, any configuration is a structure over $\mathcal{L}_Q$ and the construction of T_τ was done in order to forbid patterns that are forbidden by τ, thus a tiling by τ is a model of T_τ.

Lemma G.2

T_τ is consistent if and only if τ can tile the plane.

Proof : It is obvious (by lemma G.1) that if τ can tile the plane, then T_τ is consistent: A tiling provides a model of T_τ.

Now suppose that T_τ has a model M. We will "unfold" M starting from a point x in it by applying the functions N, S, E, W that will give us any point in $\mathbb{Z}^2$. We can define a configuration c, such that $c(0,0)$ has the "state" of x, and $c(i,j)$ has the "state" of $E^i(N^j((x)))$. This configuration is a tiling: As M is a model of T_τ, no forbidden pattern can appear. Therefore, from any model of T_τ, we can obtain a tiling of the plane by τ, which finishes the proof. ∎

Remark. T_τ has a model if and only if T_τ has an infinite model.

We can force all models to be infinite by adding (infinitely many) axioms that will ensure this property. The theory may however not be finitely axiomatisable anymore.

Note however that if a tileset does not admit any periodic tiling, no finite models can appear. Moreover, if a tileset does not admit any tiling with at least one direction of periodicity, then all models are only union of copies of $\mathbb{Z}^2$. That is, no degenerate torus or cylinder may appear.

Lemma G.3

T_τ has a finite model if and only if τ can tile periodically the plane.

Proof : Consider a periodic tiling of period p, we "fold" it into $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and obtain a model of T_τ since the cell at position $(x+p, y)$ will have the same state as the one at (x, y) or $(x, y+p)$.

If we have a finite model, we unfold it the same way as in Lemma G.2. It is easy to see that we obtain this way a periodic tiling. ∎

Most of these results can be generalized to tilings of $\mathbb{R}^2$ using "patches" as tiles and we still get the same translation from tileset and tilings into model theory [Oge04].

G.4.1 Isomorphism

One of the first properties of models of a given theory one has to study is the isomorphism of models. The translation of this property as properties of tilings is quite straightforward:

Lemma G.4

Two configurations are equal up to shift if and only if they are isomorphic as structures on the language $\mathcal{L}_Q$.

Proof : $\Rightarrow$: Let x, y be two configurations equal up to shift and σ be a shift of vector (i, j) such that $x = \sigma(y)$. Then σ is an isomorphism from x to y.

$\Leftarrow$: Let Θ be the isomorphism and a and b two points of x and y such that $\Theta(a) = b$. Then $E^i N^j(a)$ has the same state as $E^i N^j(b)$, as the predicate $P_q(E^i N^j(x))$ has the same value in a and b since Θ is an isomorphism. ∎

G.4.2 Elementary equivalence

Another model theoretic property that translates to tilings is the elementary equivalence. We recall that two structures are elementary equivalent if and only if they satisfy the same formulae (that is have the same theory)

Lemma G.5 (*[Poi80, Oge04]*)

Two configurations x and y are locally isomorphic if and only if they are elementary equivalent as structures over $\mathcal{L}_Q$.

Proof : We will consider for the proof x and y as structures over the language without functions, *i.e.* we replace in $\mathcal{L}_Q$ the functions N, S, E, W by functional predicates N', S', E', W', that is $N'(x, y) \Leftrightarrow N(x) = y$.

$\Leftarrow$: One can express the apparition of the pattern M by a first order formula like in Definition G.8: $\exists x, \varphi_M(x)$. Therefore, as any formula valid in one structure is valid in the other one, any pattern that appears in one tiling appears in the other one. This proves that if the structures are elementary equivalent then the tilings are locally isomorphic.

$\Rightarrow$: This proof is rather technical and is given in [Oge04] using Hanf locality lemma [Han65](lemma 2.3). Hanf locality lemma states that for two structures, if the spheres (using the relational distance) all contain finitely many points (what is always true in our case), and if both stuctures have either the same finite number of different spheres or both have an infinite number, then the two structures are elementary equivalent. Hanf locality lemma can be proved using a back and forth method, or an Ehrenfeucht-FraÃ¯ssÃ© game.

In our case, the spheres represent the patterns: Consider a point x and all the points at relational distance at most n, since our language contains only binary predicates and that they represent the functions N, S, E, W, the relational distance is nothing but the L_1 distance (or Manhattan distance or also Taxicab Metric) on $\mathbb{Z}^2$. Therefore the sphere at point x of radius n is the pattern defined on $B_1(x, n)$.

Both configurations x and y have the same patterns thus if a pattern appears only a finite number of times in x, it appears the same number of times in both configurations. As a consequence, Hanf lemma applies: x and y, having the same patterns, have the same theory. ∎

This theorem allows us to get an equivalence between the completeness of T_τ and a property of the tileset τ:

Theorem G.6

A tileset τ can produce only one tiling up to local isomorphism if and only if $T_\tau^\infty = T_\tau \cup \{\forall x, E^n(N^m(x)) \neq x | m, n \in \mathbb{Z}\}$ is complete.

Note that the additional axioms ensure that no model of T_τ is skewed, that is all models of T_τ are based on $\mathbb{Z}^2$ or disconnected copies of $\mathbb{Z}^2$.

There is no need indeed for these additional axioms if we can ensure that the only (up to local isomorphism) tiling by τ is actually strictly aperiodic (that is has no vector of periodicity).

Before going on the proof of Theorem G.6, we first need an extra lemma on tilings:

Lemma G.7

If all the tilings produced by a tileset τ are locally isomorphic then every pattern that appears in a tiling appears infinitely many times in it.

Proof : Consider a tiling x and suppose that there exists a pattern that appears only finitely many times. By compactness, we can extract a tiling that does not contain this pattern since we can have arbitrary large patterns that do not overlap with it. The extracted tiling that does not contain this pattern will thus not have the same patterns as x. ■

Proof (of the theorem): $\Rightarrow$: We prove that any two models of T_τ^∞ are elementary equivalent. This is already true for models that are tilings (Lemma G.5) but we still have to prove it for arbitrary models. Consider two models M and M' of T_τ^∞, they are made of disconnected copies of tilings; all patterns that appear in a tiling appear infinitely many times therefore all the spheres that appear in M or M' appear infinitely many times. Thus the hypothesis of Hanf locality lemma hold, so $M \equiv M'$. Therefore T_τ^∞ is complete.

$\Leftarrow$: If T_τ^∞ is complete, for any pattern M, the formula $\exists x, \varphi_M(x)$ is either valid in any model or false in any model, therefore any two tilings contain exactly the same patterns, thus τ can produce only one tiling up to local isomorphism. ■

Corollary G.8

If a tileset τ can produce only one tiling up to local isomorphism then the appeareance of any pattern is a decidable problem.

This is a corollary of Theorem G.6 that we express here without any model theoretic language: τ can produce only one tiling up to local isomorphism thus T_τ^∞ is complete. Given a pattern M, one can enumerate the valid proofs in T_τ^∞ and stop when either a proof of $\exists x, \varphi_M(x)$ or of $\neg\exists x, \varphi_M(x)$ is found; and such a proof will be found since T_τ^∞ is complete.

On compactness

With all those results one could try to prove some results about tilings in an elegant and short way using model theoretic arguments. Take for example the fact that any tileset that produces only periodic tilings can produce only finitely many of them [I7]. This can be reformulated as "if a tileset can produce tilings with arbitrarily large periods then it can produce one that is not periodic". It is easy to write a formula ϕ_n that expresses that there is a tiling with no period lower than n. If a tileset can produce tilings with arbitrarily large periods then it has a model verifying any finite set of such formulae, thus by compactness it has a model that verifies all these formulae, e.g., it has a model that has no period. However, we can not conclude that the tileset can produce a tiling with no period. Indeed this model we obtain by compactness will certainly consist of a copy of each periodic tiling : As we have tilings of arbitrary large period, there is no common period for all these tilings, so that our model indeed does not have a period.

We would like to be able to use the compactness theorem of the first order logic but within the domain of connected models. However as said earlier, many classical theorems of first order logic will not hold. See [Hou08] for some possible solutions.

G.4.3 Applying the results to model theory

A finitely axiomatizable, complete and superstable theory has been exhibited with these methods of translating tilesets into first order theories. This has historically been done by Makowsky [Mak74] to prove that these three properties of theories are not incompatible and then explained in a more detailed way by Poizat [Poi80].

The idea is quite simple: take τ an aperiodic tileset that produces only one tiling up to local isomorphism; for example the one used by Berger to prove the undecidability of the domino problem [Ber64]. Transform it in a first order theory as explained in Section G.3 to obtain a finitely axiomatized theory T_τ. Since Berger proved that his tileset can not produce any tiling with a vector of periodicity, Theorem G.6 holds without any need to add more axioms to ensure that the models are infinite by Lemma G.3; therefore T_τ is complete and finitely axiomatizable.

We can then prove that the theory is superstable. This definition has to do with how many types there are in the theory, or more simply, with how many tilings we can produce.

It has been proven that this tileset can produce $2^{\aleph_0}$ different tilings [Dur99, I7], therefore $2^{\aleph_0}$ countable models; Those models are not isomorphic because there is only a countable number of shifts. Furthermore, there is no skewed models, so that all models of this theory are then easy to give : they consists of some copies of these $2^{\aleph_0}$ different tilings, that is we have to say for each tiling how many times it appears. This shows that the theory is not ω-stable, but superstable.

G.5 Conclusion

We have seen along this paper the tight links between tilings and logic, especially between tilings properties and model theoretical properties of their interpretation. Tilings have then provided interesting examples of theories [Poi80] as well as a good framework in which to study properties of classes of formulaes [BGG01].

Some links still remain unexplored and might lead to interesting results. As an exemple, the Cantor-Bendixson rank [Kur66] introduced in [I7] has been motivated by the study of a notion of rank for finitely generated structures of universal theories in [Hou08].

Annexe H

SUBSHIFTS AS MODELS FOR MSO LOGIC

Dans cet article coécrit avec Guillaume Theyssier, on montre comment certains fragments de la logique monadique correspondent aux espaces de pavages (voir le chapitre 4). Il s'agit ici de la version longue soumise, et non pas de l'article de conférences [I6].

Abstract

We study the Monadic Second Order (MSO) Hierarchy over colourings of the discrete plane, and draw links between classes of formula and classes of subshifts. We give a characterization of existential MSO in terms of projections of tilings, and of universal sentences in terms of combinations of "pattern counting" subshifts. Conversely, we characterise logic fragments corresponding to various classes of subshifts (subshifts of finite type, sofic subshifts, all subshifts). Finally, we show by a separation result how the situation here is different from the case of tiling pictures studied earlier by Giammarresi et al.

H.1 Introduction

There is a close connection between words and monadic second-order (MSO) logic. Büchi and Elgot proved for finite words that MSO-formulas correspond exactly to regular languages. This relationship was developed for other classes of labeled graphs; trees or infinite words enjoy a similar connection. See [Tho97, Mat98b] for a survey of existing results. Colorings of the entire plane, i.e tilings, represent a natural generalization of biinfinite words to higher dimensions, and as such enjoy similar properties. We plan to study in this paper tilings for the point of view of monadic second-order logic.

From a computer science point of view, tilings and more generally subshifts are the underlying objects of several computing models including cellular automata [Del11, BT10, BT09], Wang tiles [LW07, LW08] and self-assembly tilings [DPR$^+$10, DLP$^+$10]. Following the recent trend to better understand such 'natural computing models', one of the motivations of the present paper is to extend towards these models the fruitful links established between languages of finite words and MSO logic.

Tilings and logic have a shared history. The introduction of tilings can be traced back to Hao Wang [Wan61], who introduced his celebrated tiles to study the (un)decidability of the $\forall\exists\forall$ fragment of first order logic. The undecidability of the domino problem by his PhD Student Berger [Ber64] lead then to the undecidability of this fragment [BGG01]. Seese [KL05, See91] used the domino problem to prove that graphs with a decidable MSO theory have a bounded tree width. Makowsky[Mak74, Poi80] used the construction by Robinson [Rob71] to give the first example of a finitely axiomatizable theory that is super-stable. More recently, Oger [Oge04] gave generalizations of classical results on tilings to locally finite relational structures. See the survey [I8] for more details.

Previously, a finite variant of tilings, called tiling pictures, was studied [GR97, GRST96]. Tiling pictures correspond to colorings of a *finite* region of the plane, this region being bordered by special '#' symbols. It is proven for this particular model that language recognized by EMSO-formulas correspond exactly to so-called finite tiling systems, i.e. projections of finite tilings.

The equivalent of finite tiling systems for infinite pictures are so-called *sofic subshifts* [Wei73]. A *sofic subshift* represents intuitively local properties and ensures that every point of the plane behaves in the same way. As a consequence, there is no general way to enforce that some specific color, say $\square$, appears at least once. Hence, some simple first-order existential formulas have no equivalent as sofic subshift (and even subshift). This is where the border of # for finite pictures play an important role: Without such a border, results on finite pictures would also stumble on this issue. See [AJM09] for similar results on finite pictures without borders.

We deal primarily in this article with subshifts. See [ATW03] for other acceptance conditions (what we called subshifts of finite type correspond to A-acceptance in this paper).

Finally, note that all decision problems in our context are non-trivial : To decide if a universal first-order formula is satisfiable (the domino problem, presented earlier) is not recursive. Worse, it is Σ_1^1-hard to decide if a tiling of the plane exists where some given color appears infinitely often [Har85, ATW03]. As a consequence, the satisfiability of MSO-formulas is at least Σ_1^1-hard.

In this paper, we will prove how various classes of formula correspond to well known classes of subshifts. Some of the results of this paper were already presented in [I6].

H.2 Symbolic Spaces and Logic

H.2.1 Configurations

Consider the discrete lattice $\mathbb{Z}^2$. For any finite set Q, a Q-configuration is a function from $\mathbb{Z}^2$ to Q. Q may be seen as a set of *colors* or *states*. An element of $\mathbb{Z}^2$ will be called a *cell*. A configuration will usually be denoted C, M or N.

Fig. H.1 shows an example of two different configurations of $\mathbb{Z}^2$ over a set Q of 5 colors. As a configuration is infinite, only a finite fragment of the configurations is represented in the figure. We choose not to represent which cell of the picture is the origin $(0, 0)$. This will indeed be of no importance as we use only translation invariant properties.

For any $z \in \mathbb{Z}^2$ we denote by σ_z the *shift* map of vector z, *i.e.* the function from Q-configurations to Q-configurations such that for all $C \in Q^{\mathbb{Z}^2}$:

$$\forall z' \in \mathbb{Z}^2,\ \sigma_z(C)(z') = C(z' - z).$$

A *pattern* is a partial configuration. A pattern $P : X \to Q$ where $X \subseteq \mathbb{Z}^2$ occurs in $C \in Q^{\mathbb{Z}^2}$ at position z_0 if

$$\forall z \in X,\ C(z_0 + z) = P(z).$$

We say that P occurs in C if it occurs at some position in C. As an example the pattern P of Fig H.2 occurs in the configuration M but not in N (or more accurately not on the finite fragment of N depicted in the figure). A finite pattern is a partial configuration of finite domain. All patterns in the following will be finite. The *language* $\mathcal{L}(C)$ of a configuration C is the set of finite patterns that occur in C. We naturally extend this notion to sets of configurations.

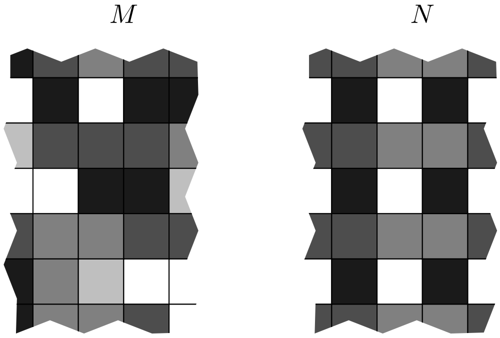


Figure H.1: Two configurations

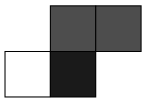


Figure H.2: A pattern P. P appears in M but presumably not in N

A *subshift* is a natural concept that captures both the notion of *uniformity* and *locality*: the only description “available” from a configuration C is the finite patterns it contains, that is $\mathcal{L}(C)$. Given a set $\mathcal{F}$ of patterns, let $X_{\mathcal{F}}$ be the set of all configurations where no patterns of $\mathcal{F}$ occurs.

$$X_{\mathcal{F}} = \{C|\mathcal{L}(C) \cap \mathcal{F} = \emptyset\}$$

$\mathcal{F}$ is usually called the set of forbidden patterns or the *forbidden language*. A set of the form $X_{\mathcal{F}}$ is called a *subshift*.

A subshift can be equivalentely defined by topology considerations. Endow the set of configurations $Q^{\mathbb{Z}^2}$ with the product topology: A sequence $(C_n)_{n\in\mathbb{N}}$ of configurations converges to a configuration C if the sequence ultimately agree with C on every $z \in \mathbb{Z}^2$. Then a subshift is a closed subset of $Q^{\mathbb{Z}^2}$ also closed by shift maps.

Example H.1

Consider the three forbidden patterns of figure H.3. The first one says that we cannot find a ■ point at the left of a ■ point. This can be interpreted as follows: every time we find a ■ point, then all the points at the right of it are also ■. With the second forbidden pattern, we deduce that every time we find a ■ point, then the entire quarter of plane on the above right of it is also filled with ■ points. The third pattern ensures us that every configuration contains at most one quarter of plane of color ■ : if it contains two such quarters of plane, then there must be a bigger quarter of plane that contains both.

Hence a typical configuration looks like A. Other possible configurations are B, C, D, E. They correspond to extremal situations where the corner of the quarter of plane is situated respectively at $(0, -\infty)$, $(-\infty, 0)$, $(-\infty, -\infty)$ et $(+\infty, +\infty)$

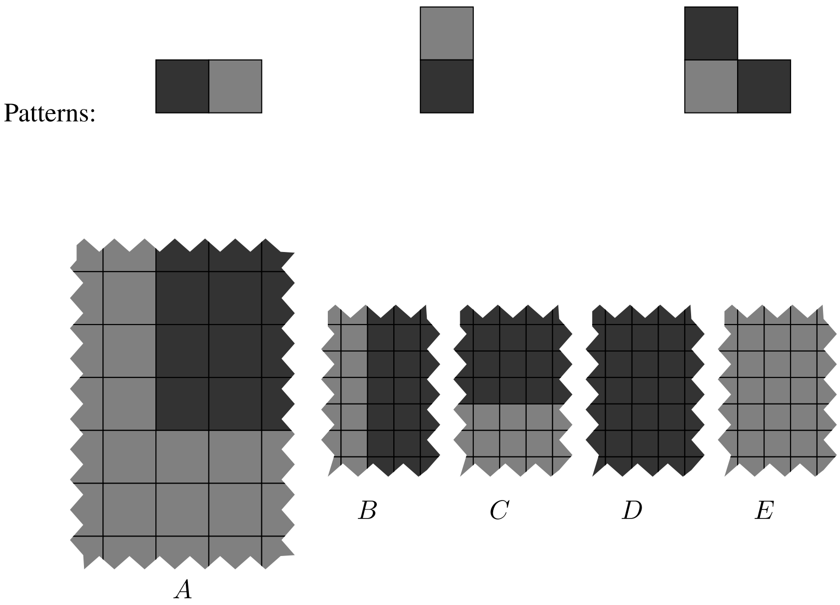


Figure H.3: A (finite) set of forbidden patterns $\mathcal{F}$ and the tilings it generates

Example H.2

> Consider the set of colors {■, □} and $\mathcal{F}$ to be the set of patterns that contains two ■ points or more.
>
> Then $X_{\mathcal{F}}$ contains configurations with at most one ■ point. Up to shift, $X_{\mathcal{F}}$ contains then two configurations: the all □-one, and one where only one point is ■ and all others are □.

A *subshift of finite type* (or *tiling*) is a subshift that can be defined via a finite set $\mathcal{F}$ of forbidden patterns: it is the set of configurations C such that no pattern in $\mathcal{F}$ occurs in C. If all patterns of $\mathcal{F}$ fit in a $n \times n$ square, this means that we only have to see a configuration through a window of size $n \times n$ to know if it is a tiling, hence the locality. Example H.1 is a subshift of finite type. It can be proven that Example H.2 is not.

Given two state sets Q_1 and Q_2, a projection is a map $\pi : Q_1 \to Q_2$. We naturally extend it to $\pi : Q_1^{\mathbb{Z}^2} \to Q_2^{\mathbb{Z}^2}$ by $\pi(C)(z) = \pi(C(z))$. A *sofic subshift* of state set Q_2 is the image by some projection π of some subshift of finite type of state set Q_1. It is also a subshift (clearly closed by shift maps, and topologically closed because projections are continuous maps on a compact space). A sofic subshift is a natural object in tiling theory, although quite never mentioned explicitly. It represents the concept of *decoration*: some of the tiles we assemble to obtain the tilings may be decorated, but we forgot the decoration when we observe the tiling.

Example H.3

Consider the following variant of Example H.1: tilings are exactly the same except that the corner of the quarter of plane in A is of a different color $\square$. It is easy to see that this variant defines a subshift of finite type X (with a few more forbidden patterns).

Now consider the following map:

$$\pi : \begin{array}{ccc} {\color{gray}\blacksquare} & \mapsto & \square \\ \blacksquare & \mapsto & \square \\ \square & \mapsto & {\color{darkgray}\blacksquare} \end{array}$$

Then B, C, D, E will become under π of color $\square$, while A will become a configuration with exactly one ${\color{darkgray}\blacksquare}$, all other points being $\square$.

As a consequence, $\pi(X)$ is exactly Example H.2. Example H.2 is thus a sofic subshift.

H.2.2 Structures

A configuration will be seen in this article as an infinite structure. The signature τ contains four unary maps North, South, East, West and a predicate P_c for each color $c \in Q$.

A configuration M will be seen as a structure $\mathfrak{M}$ in the following way:

– The elements of $\mathfrak{M}$ are the points of $\mathbb{Z}^2$.
– North is interpreted by $\mathsf{North}^{\mathfrak{M}}((x,y)) = (x, y+1)$. $\mathsf{East}^{\mathfrak{M}}$, $\mathsf{South}^{\mathfrak{M}}$ and $\mathsf{West}^{\mathfrak{M}}$ are interpreted similarly
– $P_c^{\mathfrak{M}}((x,y))$ is true if and only if the point at coordinate (x,y) is of color c, that is if $M(x,y) = c$.

As an example, the configuration M of Fig. H.1 has three consecutive cells with the color ${\color{darkgray}\blacksquare}$. That is, the following formula is true:

$$\mathfrak{M} \models \exists z, P_{{\color{darkgray}\blacksquare}}(z) \wedge P_{{\color{darkgray}\blacksquare}}(\mathsf{East}(z)) \wedge P_{{\color{darkgray}\blacksquare}}(\mathsf{East}(\mathsf{East}(z)))$$

As another example, the following formula states that the configuration has a vertical period of 2 (the color in the cell (x,y) is the same as the color in the cell $(x, y+2)$). The formula is false in the structure $\mathfrak{M}$ and true in the structure $\mathfrak{N}$ (if the reader chose to color the cells of N not shown in the picture correctly):

$$\forall z, \begin{cases} P_{{\color{gray}\blacksquare}}(z) \implies P_{{\color{gray}\blacksquare}}(\mathsf{North}(\mathsf{North}(z))) \\ P_{\square}(z) \implies P_{\square}(\mathsf{North}(\mathsf{North}(z))) \\ P_{\blacksquare}(z) \implies P_{\blacksquare}(\mathsf{North}(\mathsf{North}(z))) \\ P_{{\color{darkgray}\blacksquare}}(z) \implies P_{{\color{darkgray}\blacksquare}}(\mathsf{North}(\mathsf{North}(z))) \\ P_{{\color{lightgray}\blacksquare}}(z) \implies P_{{\color{lightgray}\blacksquare}}(\mathsf{North}(\mathsf{North}(z))) \end{cases}$$

Remark. The choice of unary function (north, south, east, west) instead of binary relations in the signature above is important because it allows a simple characterization of important classes of subshifts (see theorem H.6 below). This particular theorem would fail with binary relations in the signature instead of unary functions. Other theorems would be still valid.

H.2.3 Monadic Second-Order Logic

This paper studies connection between subshifts (seen as structures as explained above) and monadic second order sentences. First order variables (x, y, z, ...) are interpreted as points of $\mathbb{Z}^2$ and (monadic) second order variables (X, Y, Z, ...) as subsets of $\mathbb{Z}^2$.

Monadic second order formulas (ϕ, ψ, ...) are defined as follows:

– a term is either a first-order variable or a function (South, North, East, West) applied to a term ;
– atomic formulas are of the form $t_1 = t_2$ or $X(t_1)$ where t_1 and t_2 are terms and X is either a second order variable or a color predicate ;
– formulas are build up from atomic formulas by means of boolean connectives and quantifiers $\exists$ and $\forall$ (which can be applied either to first-order variables or second order variables).

A formula is *closed* if no variable occurs free in it. A formula is FO if no second-order quantifier occurs in it. A formula is EMSO if it is of the form

$$\exists X_1, \ldots, \exists X_n, \phi(X)$$

where ϕ is FO. Given a formula $\phi(X_1, \ldots, X_n)$ with no free first-order variable and having only $X_1, \ldots, X_n$ as free second-order variables, a configuration M together with subsets $E_1, \ldots, E_n$ is a model of $\phi(X_1, \ldots, X_n)$, denoted

$$(M, E_1, \ldots, E_n) \models \phi(X_1, \ldots, X_n),$$

if ϕ is satisfied (in the usual sense) when M is interpreted as $\mathfrak{M}$ (see previous section) and E_i interprets X_i.

H.2.4 Definability

This paper studies the following problems: Given a formula ϕ of some logic, what can be said of the configurations that satisfy ϕ? Conversely, given a subshift, what kind of formula can characterise it?

Definition H.4

A set S of Q-configurations is defined by ϕ if

$$S = \left\{ M \in Q^{\mathbb{Z}^2} \middle| \mathfrak{M} \models \phi \right\}$$

Two formulas ϕ and ϕ' are equivalent iff they define the same set of configurations.

A set S is $\mathcal{C}$-definable if it is defined by a formula $\phi \in \mathcal{C}$.

It is easy to see that Example H.1 is defined by the formula

$$\phi : \begin{cases} \forall x, \neg \left(P_{\blacksquare}(x) \wedge P_{\textcolor{gray}{\blacksquare}}(\mathsf{East}(x)) \right) \\ \forall x, \neg \left(P_{\blacksquare}(x) \wedge P_{\textcolor{gray}{\blacksquare}}(\mathsf{North}(x)) \right) \\ \forall x, \neg \left(P_{\textcolor{gray}{\blacksquare}}(x) \wedge P_{\blacksquare}(\mathsf{East}(x)) \wedge P_{\textcolor{gray}{\blacksquare}}(\mathsf{North}(x)) \right) \end{cases}$$

or equivalently by the formula

$$\phi' : \forall x, P_{\blacksquare}(x) \iff \left(P_{\blacksquare}(\mathsf{East}(x)) \wedge P_{\blacksquare}(\mathsf{North}(x)) \right)$$

We will see some variants of formula ϕ' appear in a few theorems below.

Example H.2 is defined by the formula

$$\psi : \forall x, y, \Big(P_{\blacksquare}(x) \wedge P_{\blacksquare}(y)\Big) \implies x = y$$

Note that a definable set is always closed by shift (a shift between 2 configurations induces an isomorphism between corresponding structures). It is not always closed: The set of $\{\blacksquare, \square\}$-configurations defined by the formula $\phi : \exists z, P_{\blacksquare}(z)$ contains all configurations except the all-white one, hence is not closed.

When we are dealing with MSO formulas, the following remark is useful: second-order quantifiers may be represented as projection operations on sets of configurations. We formalize now this notion.

If $\pi : Q_1 \mapsto Q_2$ is a projection and S is a set of Q_1-configurations, we define the two following operators:

$$\begin{aligned} E(\pi)(S) &= \Big\{M \in (Q_2)^{\mathbb{Z}^2} \Big| \exists N \in (Q_1)^{\mathbb{Z}^2}, \pi(N) = M \wedge N \in S\Big\} \\ A(\pi)(S) &= \Big\{M \in (Q_2)^{\mathbb{Z}^2} \Big| \forall N \in (Q_1)^{\mathbb{Z}^2}, \pi(N) = M \implies N \in S\Big\} \end{aligned}$$

Note that A is a dual of E, that is $A(\pi)(S) = {}^cE(\pi)({}^cS)$ where ^{c} represents complementation.

Proposition H.1

- A set S of Q-configurations is EMSO-definable if and only if there exists a set S' of Q' configurations and a map $\pi : Q' \mapsto Q$ such that $S = E(\pi)(S')$ and S' is FO-definable.
- The class of MSO-definable sets is the closure of the class of FO-definable sets by the operators E and A.

Proof (Sketch): Second item is a straightforward reformulation of the prenex normal form of MSO using operators E and A. We prove here only the first item.

- Let $\phi = \exists X, \psi$ be a EMSO formula that defines a set S of Q-configurations. Let $Q' = Q \times \{0, 1\}$ and π be the canonical projection from Q' to Q.
  Consider the formula ψ' obtained from ψ by replacing $X(t)$ by $\vee_{c \in Q} P_{(c,1)}(t)$ and $P_c(t)$ by $P_{(c,0)}(t) \vee P_{(c,1)}(t)$.
  Let S' be a set of Q' configurations defined by ψ'. Then is it clear that $S = E(\pi)(S')$.
  The generalization to more than one existential quantifier is straightforward.
- Let $S = E(\pi)(S')$ be a set of Q configurations, and S' FO-definable by the formula ϕ. Denote by $c_1 \dots c_n$ the elements of Q'. Consider the formula ϕ' obtained from ϕ where each P_{c_i} is replaced by X_i. Let

$$\psi = \exists X_1, \dots, \exists X_n, \left\{ \begin{array}{l} \forall z, \vee_i X_i(z) \\ \forall z, \wedge_{i \neq j} (\neg X_i(z) \vee \neg X_j(z)) \\ \forall z, \wedge_i \big(X_i z \implies P_{\pi(c_i)}(z)\big) \\ \phi' \end{array} \right.$$

  Then ψ defines S. Note that the formula ψ constructed above is of the special form $\exists X_1, \dots, \exists X_n (\forall z, \psi'(z)) \wedge \phi'$. This will be important later. ■

Second-order quantifications will then be regarded in this paper either as projections operators or sets quantifiers.

H.3 Hanf Locality Lemma and EMSO

The first-order logic has a property that makes it suitable to deal with tilings and configurations: it is local. This is illustrated by Hanf's lemma [Han65, EF95, Lib04]. A square pattern of radius n is a pattern of domain $[-n, n] \times [n, n]$.

Definition H.5

Two Q-configurations M and N are (n, k)-equivalent if for each Q-square pattern P of radius n:

- If P appears in M at most k times, then P appears the exact same number of times in M and in N
- If P appears in M more than k times, then P appears in N more than k times

This notion is indeed an equivalence relation. Given n and k, it is clear that there is only finitely many equivalence classes for this relation.

Contrary to definition H.5 above, Hanf's original formalism doesn't use square shapes (balls for the $\|\cdot\|_\infty$ norm) but lozenges (balls for the $\|\cdot\|_1$ norm). It makes essentially no difference and the Hanf's local lemma can be reformulated in our context as follows (proofs using formalism of definition H.5 appear in [GRST96]).

Theorem H.2

For every FO formula ϕ, there exists (n, k) such that

$$\text{if } M \text{ and } N \text{ are } (n, k) \text{ equivalent, then } \mathfrak{M} \models \phi \iff \mathfrak{N} \models \phi$$

Corollary H.3

Every FO-definable set is a (finite) union of some (n, k)-equivalence classes.

This is theorem 3.3 in [GRST96], stated for finite configurations. Lemma 3.5 in the same paper gives a proof of Hanf's Local Lemma in our context.

Given (P, k) we consider the set $S_{=k}(P)$ of all configurations such that the pattern P occurs exactly k times (k may be taken equal to 0). The set $S_{\geq k}(P)$ is the set of all configurations such that the pattern P occurs k times or more.

We may rephrase the preceding corollary as:

Corollary H.4

Every FO-definable set is a positive combination (i.e. unions and intersections) of some $S_{=k}(P)$ and some $S_{\geq k}(P)$

Theorem H.5

Every EMSO-definable set can be defined by a formula ϕ of the form:

$$\exists X_1, \ldots, \exists X_n, \big(\forall z_1,\ \phi_1(z_1, X_1, \ldots, X_n)\big) \tag{H.1}$$
$$\wedge \big(\exists z_1, \ldots, \exists z_p,\ \phi_2(z_1 \ldots z_p, X_1, \ldots, X_n)\big), \tag{H.2}$$

where ϕ_1 and ϕ_2 are quantifier-free formulas.

See [Tho97, Corollary 4.1] or [Tho91, Corollary 4.2] for a similar result. This result is an easy consequence of [SB99, Theorem 3.2] (see also the corrigendum). We include here a full proof.

Proof : Let $\mathcal{C}$ be the set of such formulas. We proceed in three steps:

– Every EMSO-definable set is the projection of a positive combination of some $S_{=k}(P)$ and $S_{\geq k}(P)$ (using prop. H.1 and the preceding corollary)
– Every $S_{=}(P,k)$ (resp. $S_{\geq}(P,k)$) is $\mathcal{C}$-definable
– $\mathcal{C}$-definable sets are closed by (finite) union, intersection and projections.

$\mathcal{C}$-definable sets are closed by projection using the equivalence of prop. H.1 in the two directions, the note at the end of the proof and some easy formula equivalences. The same goes for intersection.

Now we prove that $\mathcal{C}$-definable sets are closed by union. The difficulty is to ensure that we use only one universal quantifier. Let ϕ and ϕ' be two $\mathcal{C}$-formulas defining sets S_1 and S_2. We can suppose that ϕ and ϕ' use the same numbers of second-order quantifiers and of first-order existential quantifiers.

Then the formula

$$\exists X, \exists X_1, \ldots, \exists X_n, \forall z_1, \left\{ \begin{array}{rcl} X(z_1) & \iff & X(\mathsf{North}(z_1)) \\ X(z_1) & \iff & X(\mathsf{East}(z_1)) \\ X(z_1) & \implies & \phi_1(z_1, X_1 \ldots X_n) \\ \neg X(z_1) & \implies & \phi_1'(z_1, X_1 \ldots X_n) \end{array} \right. \tag{H.3}$$

$$\wedge\, \exists z_1, \ldots, \exists z_p \bigvee \begin{array}{c} X(z_1) \wedge \phi_2(z_1 \ldots z_p, X_1 \ldots X_n) \\ \neg X(z_1) \wedge \phi_2'(z_1 \ldots z_p, X_1 \ldots X_n) \end{array} \tag{H.4}$$

defines $S_1 \cup S_2$ (the disjunction is obtained through variable X which is forced to represent either the empty set or the whole plane $\mathbb{Z}^2$).

It is now sufficient to prove that a $S_{=k}(P)$ set (resp. a $S_{\geq k}(P)$ set) is definable by a $\mathcal{C}$-formula. Let $\phi_P(z)$ be the quantifier-free formula such that $\phi_P(z)$ is true if and only if P appears at position z.

Then $S_{=k}(P)$ is definable by

$$\exists X_1 \ldots \exists X_k \exists A_1, \ldots, \exists A_k, \forall x \left\{ \begin{array}{lcl} \wedge_i A_i(x) & \iff & [A_i(\mathsf{North}(x)) \wedge A_i(\mathsf{East}(x))] \\ \wedge_i X_i(x) & \iff & [A_i(x) \wedge \neg A_i(\mathsf{South}(x)) \wedge \neg A_i(\mathsf{West}(x))] \\ \wedge_{i \neq j} X_i(x) & \implies & \neg X_j(x) \\ (\vee_i X_i(x)) & \iff & \phi_P(x) \end{array} \right. \tag{H.5}$$

$$\wedge\, \exists z_1, \ldots, \exists z_k, X_1(z_1) \wedge \cdots \wedge X_k(z_k) \tag{H.6}$$

The formula ensures indeed that A_i represents a quarter of the plane, X_i being a singleton representing the corner of that plane. If $k = 0$ this becomes $\forall x, \neg\phi_P(x)$. To obtain a formula for $S_{\geq k}(P)$, change the last $\iff$ to a $\implies$ in the formula. ∎

H.4 Characterization of Subshifts of Finite Type and Sofic Subshifts

H.4.1 Subshifts of Finite Type

We start by a characterization of subshifts of finite type (SFTs, i.e tilings). The problem with SFTs is that they are closed neither by projection nor by union: the 'even shift' is the projection of a SFT but is not itself a SFT (see [LM95]) and if $\mathcal{F}_1 = \{\blacksquare\square\}$ and $\mathcal{F}_2 = \{\square\blacksquare\}$ then the union $X_{\mathcal{F}_1} \cup X_{\mathcal{F}_2}$ is not a SFT. As a consequence, the class of formulas corresponding to SFTs is not very interesting:

Theorem H.6

A set of configurations is a SFT if and only if it is defined by a formula of the form

$$\forall z,\ \psi(z)$$

where ψ is quantifier-free.

Note that there is only one quantifier in this formula. Formulas with more than one universal quantifier do not always correspond to SFT: This is due to SFTs not being closed by union.

Proof : Let $P_1 \dots P_n$ be patterns. To each P_i we associate the quantifier-free formula $\phi_{P_i}(z)$ which is true if and only if P_i appears at the position z. Then the subshifts that forbids patterns $P_1 \dots P_n$ is defined by the formula:

$$\forall z, \neg\phi_{P_1(z)} \wedge \dots \wedge \neg\phi_{P_n(z)}$$

Conversely, let ψ be a quantifier-free formula. Each term t_i in ψ is of the form $f_i(z)$ where f_i is some combination of the functions North, South, East and West, each f_i thus representing somehow some vector z_i ($f_i(z) = z + z_i$). Let Z be the collection of all vectors z_i that appear in the formula ψ. Now the fact that ψ is true at the position z only depends on the colors of the configurations in points $(z + z_1), \dots, (z + z_n)$, i.e. on the *pattern* of domain Z that occurs at position z. Let $\mathcal{P}$ be the set of patterns of domain Z that makes ψ false. Then the set S defined by ψ is the set of configurations where no patterns in $\mathcal{P}$ occurs, hence a SFT. ■

H.4.2 Universal sentences

Due to the way subshifts are defined, universal quantifiers play an important role. We now ask the following question: what are the sets defined by universal formulas? First the following lemma shows that we can restrict to first-order when considering universal formulas.

Lemma H.7

Any universal MSO formula is equivalent to a first-order universal formula.

Proof : A universal formula is equivalent (through permutation of universal quantifiers) to a formula of the form

$$\forall x_1, \dots, x_p, \forall X_1, \dots, X_n,\ \Phi(X_1, \dots, X_n, x_1, \dots, x_p)$$

where Φ is quantifier-free. Consider the formula

$$\psi(X_1, \dots, X_{n-1}, x_1, \dots, x_p) \equiv \forall X_n, \Phi(X_1, \dots, X_n, x_1, \dots, x_p)$$

Let $\{t_1, \dots, t_k\}$ be the set of terms t such that $X_n(t)$ occurs in Φ. The idea is that the truth value of $\Phi(X_1, \dots, X_n, x_1, \dots, x_p)$ depends only on the value of X_n at positions represented by the (t_i). Depending on interpretations of the variables (x_i), interpretations of the terms (t_i) may be equal or not. We say an assignment $\rho : \{1, \dots, k\} \to \{0, 1\}$ is *sound* if $t_i = t_j \implies \rho(i) = \rho(j)$. Denote by $\phi_\rho(x_1, \dots, x_p)$ the quantifier-free formula expressing this condition:

$$\phi_\rho(x_1, \dots, x_p) \equiv \bigwedge_{\{(i,j):\rho(i)\neq\rho(j)\}} t_j \neq t_j.$$

Let ψ_ρ denote the formula $\Phi\big[X_n(t_i) \leftarrow \rho(i)\big]$ obtained from Φ be replacing each occurrence of $X_n(t_i)$ by the truth value $\rho(i)$ and this for each $i \in \{1, \dots, k\}$. For any fixed $x_1, \dots, x_p$, the truth value of $\forall X_n \Phi(X_1, \dots, X_n, x_1, \dots, x_p)$ is the same as the truth value of the conjunction of formulas ψ_ρ for all sound ρ. Hence, we get that $\psi(X_1, \dots, X_{n-1}, x_1, \dots, x_p)$ is equivalent to the following quantifier-free formula:

$$\bigwedge_{\rho:\{1,\dots,k\}\to\{0,1\}} \phi_\rho \implies \psi_\rho.$$

We can eliminate this way second order universal quantifiers one by one and the lemma follows. ■

For the rest of this section we focus on first-order universal formulas. The real difficulty is to treat the equality predicate (=). Without the equality (more precisely if all predicates and functions are only unary) any first-order universal formula is equivalent to a conjonction of formulas with only one quantifier and theorem H.6 applies. The equality predicate intertwines the variables and makes thing a bit harder to prove. The reader might for example try to understand what the following formula exactly means:

$$\forall x, y, \Big(P_{\blacksquare}(x) \wedge P_{\blacksquare}(\mathsf{East}(y))\Big) \implies x = y$$

To understand it, we will prove an analog of Hanf's Lemma for universal sentences.

Definition H.6

Let (n, k) be integers, and M, N two Q-configurations. We say that $M \geq_{n,k} N$ if for each Q-square pattern P of radius less than n:
- If P appears in M exactly p times and $p \leq k$, then P appears at most p times in N

Note that M and N are (n, k) equivalent if and only if $M \geq_{n,k} N$ and $N \geq_{n,k} M$.

Theorem H.8

For every universal formula ϕ there exists (n, k) such that if $M \geq_{n,k} N$, then $\mathfrak{M} \models \phi \implies \mathfrak{N} \models \phi$

Compare with definition H.5 and theorem H.2. Note that Gaifman's Theorem (a more refined version of Hanf's lemma) was generalized in [GW04] to existential sentences. We may use this result to obtain ours. We give below a complete direct proof.

Proof : We will translate the usual proof of Hanf's Local Lemma into our special case. We will try as much as possible to use the same notations as [EF95, sec. 2.4].

We first change the vocabulary and consider that East, West, North, South are binary predicates rather than functions. Note that every universal formula will remain a universal formulas, albeit with more quantifiers.

Let introduce some notations. Let $S(r, a)$ be the set of all points at distance at most r of a. That is $S(r, a) = \{x : |x - a| \leq r\}$ where $|\cdot|$ is the Manhattan distance. Note that $S(r, a)$ contains $e_r = 2r^2 + 2r + 1$ points. Let $S(r, a_1 \dots a_p) = \cup_i S(r, a_i)$.

Let M and N be two Q-configurations. We say that $a_1 \dots a_p \in (\mathbb{Z}^2)^p$ and $b_1 \dots b_p \in (\mathbb{Z}^2)^p$ are k-isomorphic if there exists a bijective map f from $S(3^k, a_1 \dots a_p)$ to $S(3^k, b_1 \dots b_p)$ that preserves the relations, that is

- x East $y \iff f(x)$ East $f(y)$
- $P_c(x) \iff P_c(f(x))$
- $f(a_i) = b_i$.

It is then clear that if $a_1 \dots a_p$ and $b_1 \dots b_p$ are 0-isomorphic, then we have $\mathfrak{M} \models \psi(a_1 \dots a_p) \iff \mathfrak{N} \models \psi(b_1 \dots b_p)$ whenever ψ is quantifier-free.

Now take a formula $\phi = \forall x_1 \dots x_n \psi(x_1 \dots x_n)$ where ψ is quantifier-free.

Let M and N such that $M \geq_{3^n, ne_{3^n}+1} N$.

We now prove by induction that

if $a_1 \dots a_p$ and $b_1 \dots b_p$ are $(n-p)$-isomorphic, then for all b_{p+1}, there exists a_{p+1} such that $a_1 \dots a_{p+1}$ and $b_1 \dots b_{p+1}$ are $(n-p-1)$-isomorphic.

- Case $p = 0$. Let $b_1 \in \mathbb{Z}^2$. Consider the pattern of radius 3^n centered around b_1 in N. This pattern appears in N, hence must appear in M at least one time. Take a_1 to be the center of this pattern.
- Case $p \mapsto p + 1$. Let $a_1 \dots a_p$ and $b_1 \dots b_p$ be $n - p$ isomorphic. Let $b_{p+1} \in \mathbb{Z}^2$.
  - Case 1: $|b_{p+1} - b_i| \leq 2 \times 3^{n-p-1}$ for some b_i.
    In this case $S(3^{n-p-1}, b_{p+1}) \subseteq S(3^{n-p}, b_i)$. Hence by taking $a_{p+1} = f^{-1}(b_{p+1})$ where f is the bijective map involved in the $n-p$ isomorphism, it is clear that $a_1 \dots a_{p+1}$ and $b_1 \dots b_{p+1}$ are $n - p - 1$ isomorphic.
  - Case 2: $\forall i, |b_{p+1} - b_i| > 2 \times 3^{n-p-1}$. In this case for every i, $S(3^{n-p-1}, b_{p+1}) \cap B(3^{n-p-1}, b_i) = \emptyset$.
    Consider the pattern P of radius 3^{n-p-1} centered around b_{p+1}.
    This pattern appears α times inside $S(2\times 3^{n-p-1}, b_1 \dots b_p)$ where $\alpha \leq pe_{2\times 3^{n-p-1}}$.
    P appears at least $\alpha + 1$ times in N and $\alpha + 1 \leq ne_{3^n} + 1$ hence must appears at least $\alpha + 1$ times in M. As it appears the same amount of time in $S(2 \times 3^{n-p-1}, b_1 \dots b_p)$ and $S(2 \times 3^{n-p-1}, a_1 \dots a_p)$ (by $n - p$ isomorphism), it must appear somewhere else, say centered in a_{p+1}. This a_{p+1} is not inside $S(3^{n-p-1}, a_1 \dots a_p)$ because otherwise it would be the center of an occurrence of pattern P inside $S(2 \times 3^{n-p-1}, a_1 \dots a_p)$. As a consequence, $a_1 \dots a_{p+1}$ and $b_1 \dots b_{p+1}$ are $n - p - 1$ isomorphic.

Now suppose that $\mathfrak{M} \models \phi$. Take $b_1 \dots b_n \in \mathbb{Z}^2$. There exists $a_1 \dots a_n$ such that $a_1 \dots a_n$ and $b_1 \dots b_n$ are 0-isomorphic. As $\mathfrak{M} \models \phi$ the quantifier-free formula $\psi(a_1 \dots a_n)$ is true in $\mathfrak{M}$. As a consequence $\psi(b_1 \dots b_n)$ is true in $\mathfrak{N}$. As this is true for all $b_1 \dots b_n$ we obtain $\mathfrak{N} \models \phi$. ■

Given (P, k) we consider the set $S_{\leq k}(P)$ of all configurations such that the pattern P occurs at most k times (k may be taken equal to 0)

Corollary H.9

A set is definable by a universal formula if and only if it is a positive combination (i.e. unions and intersections) of some $S_{\leq k}(P)$.

This corollary should be compared to corollary H.4.

Proof : Let $\mathcal{C}$ be the class of all universal formulas. It is clear that the set of $\mathcal{C}$-defined formulas is closed under intersection and unions.

Now $S_{\leq k}(P)$ is defined by

$$\forall x_1 \dots x_{k+1}, \phi_P(x_1) \wedge \dots \wedge \phi_P(x_{k+1}) \implies \bigvee_{i \neq j} x_i = x_j$$

For $k = 0$, this becomes $\forall x, \neg\phi_P(x)$. Hence, every positive combination of some $S_{\leq k}(P)$ is $\mathcal{C}$-definable.

Conversely, let ϕ be a universal formula and S the set it defines. Let (n, k) be as in the theorem.

For each configuration $M \in S$ and P a pattern of radius less than or equal to n, denote $\phi_M(P)$ the number of times P appears in M with the convention than $\phi_M(P) = \infty$ if P appears more than k times in M.

Consider the set

$$S_M = \bigcap_{\substack{P \mid \phi_M(P) \neq \infty, \\ \text{radius}(P) \leq n}} S_{\leq \phi_M(P)}(P)$$

From the hypothesis on (n, k), we have $S_M \subseteq S$. It is then easy to see that $S = \cup_M S_M$ where the union is actually finite (two configurations that are (n, k)-equivalent give the same S_M). ■

H.4.3 Sofic subshifts

Recall that sofic subshifts are projections of SFTs. Using the previous corollary, we are now able to give a characterisation of sofic subshifts:

Theorem H.10

A set S is a sofic subshift if and only if it is definable by a formula of the form

$$\exists X_1, \exists X_2 \dots, \exists X_n, \forall z_1, \dots, \forall z_p, \ \psi(X_1, \dots, X_n, z_1 \dots z_p)$$

where ψ is quantifier-free. Moreover, any such formula is equivalent to a formula of the same form but with a single universal quantifier ($p = 1$).

See [I6] for a different proof that eliminates equality predicates one by one.

Proof : Let $\mathcal{C}$ be the clas of all formulas of the form

$$\exists X_1, \dots, \exists X_n, \forall z \psi(X_1, \dots, X_n, z)$$

where ψ is quantifier-free. With the help of theorem H.6 and proposition H.1, is is quite clear that $\mathcal{C}$-defined sets are exactly sofic subshifts.

Let $\mathcal{D}$ be the class of all formulas of the form

$$\exists X_1, \ldots, \exists X_n, \forall z_1 \ldots z_p \psi(X_1, \ldots, X_n, z_1 \ldots z_p)$$

where ψ is quantifier-free. The previous remark states that sofic subshifts are $\mathcal{D}$-defined.

Now we prove that $\mathcal{D}$-defined sets are sofic subshifts. Using (the proof of) proposition H.1, and the fact that sofic subshifts are closed under projection, it is sufficient to prove that universal formulas define sofic subshifts. Using corollary H.9 and the fact that sofic subshifts are closed under union and projections, it is sufficient to prove that every $S_{\leq k}(P)$ is sofic.

Now $S_{\leq k}(P)$ is defined by

$$\phi : \exists S_1 \ldots S_k \begin{cases} \Psi_i \\ \forall x, \vee_i S_i(x) \iff \phi_P(x) \end{cases}$$

where Ψ_i expresses that S_i has at most one element and is defined as follows:

$$\Psi_i \stackrel{def}{=} \exists A, \forall x \begin{cases} A(x) \iff A(\mathsf{North}(x)) \wedge A(\mathsf{East}(x)) \\ S_i(x) \iff A(x) \wedge \neg A(\mathsf{South}(x)) \wedge \neg A(\mathsf{West}(x)) \end{cases}$$

Now with some light rewriting we can transform ϕ into a formula of the class $\mathcal{C}$, which proves that $S_{\leq k}(P)$ is $\mathcal{C}$-definable, hence sofic. ∎

H.5 (E)MSO-definable subshifts

H.5.1 Separation result

Theorems H.5 and H.10 above suggest that EMSO-definable subshifts are not necessarily sofic. We will show in this section that the set of EMSO-definable subshifts is indeed strictly larger than the set of sofic subshifts. The proof is based on the analysis of the computational complexity of forbidden languages (the complement of the set of patterns occuring in the considered subshift). It is well-known that any sofic subshift X has a recursively enumerable forbidden language: first, with a straightforward backtracking algorithm, we can recursively enumerate all patterns that do not occur in a given SFT Y; second, if X is the projection of Y, we can recursively enumerate all patterns P such that all patterns Q that projects onto P are forbidden in Y. The following theorem shows that the forbidden language of an MSO-definable subshift can be arbitrarily high in the arithmetical hierarchy.

This is not surprising since arbitrary Turing computation can be defined via first order formulas (using tilesets) and second order quantifiers can be used to simulate quantification of the arithmetical hierarchy. However, some care must be taken to ensure that the set of configurations obtained is a subshift.

Theorem H.11

Let E be an arithmetical set. Then there is an MSO-definable subshift with forbidden language $\mathcal{F}$ such that E reduces to $\mathcal{F}$ (for many-one reduction).

Proof sketch: Suppose that the complement of E is defined as the set of integers m such that:

$$\exists x_1, \forall x_2, \ldots, \exists/\forall x_n, R(m, x_1, \ldots, x_n)$$

where R is a recursive relation. We first build a formula ϕ defining the set of configurations representing a successful computation of R on some input $m, x_1, \ldots, x_n$. Consider 3 colors c_l, c and c_r and additional second order variables $X_1, \ldots, X_n$ and $S_1, \ldots, S_n$. The input $(m, x_1, \ldots, x_n)$ to the computation is encoded in unary on an horizontal segment using colors c_l and c_r and variables S_i as separators, precisely: first an occurrence of c_l then m occurrences of c, then an occurrence of c_r and, for each successive $1 \leq i \leq n$, x_i positions in X_i before a position of S_i. Let ϕ_1 be the FO formula expressing the following:

1. there is exactly 1 occurrence of c_l and the same for c_r and all S_i are singletons;
2. starting from an occurrence c_l and going east until reaching S_n, the only possible successions of states are those forming a valid input as explained above.

Now, the computation of R on any input encoded as above can be simulated via tiling constraints in the usual way. Consider sufficiently many new second order variables $Y_1, \ldots, Y_p$ to handle the computation and let ϕ_2 be the FO formula expressing that:

1. a valid computation starts at the north of an occurrence of c_l;
2. there is exactly one occurrence of the halting state (represented by some Y_i) in the whole configuration.

We define ϕ by:

$$\exists X_1, \forall X_2, \ldots, \exists/\forall X_n, \exists S_1, \ldots, \exists S_n, \exists Y_1, \ldots, \exists Y_p, \phi_1 \wedge \phi_2. \tag{H.7}$$

Finally let ψ be the following FO formula: $(\forall z, \neg P_{c_l}) \vee (\forall z, \neg P_{c_r})$. Let X be the set defined by $\phi \vee \psi$. By construction, a finite (unidimensional) pattern of the form $c_l c^m c_r$ appears in some configuration of X if and only if $m \notin E$. Therefore E is many-one reducible to the forbidden language of X.

To conclude the proof it is sufficient to check that X is closed. To see this, consider a sequence $(C_n)_n$ of configurations of X converging to some configuration C. C has at most one occurrence of c_l and one occurrence of c_r. If one of these two states does not occur in C then $C \in X$ since ψ is verified. If, conversely, both c_l and c_r occur (once each) then any pattern containing both occurrences also occurs in some configuration C_n verifying ϕ. But ϕ is such that any modification outside the segment between c_l and c_r in C_n does not change the fact that ϕ is satisfied provided no new c_l and c_r colors are added. Therefore ϕ is also satisfied by C and $C \in X$. ■

The theorem gives the claimed separation result for subshifts of EMSO.

Corollary H.12

There are EMSO-definable subshifts which are not sofic.

Proof : In the previous theorem, choose E, to be the complement of the set of integers m for which there is x such that machine m halts on empty input in less than x steps. E is not recursively enumerable and, using the construction of the proof above, it is reducible to the forbidden language of an EMSO-definable subshift. ■

H.5.2 Subshifts and patterns

In the previous section we proved that there exists a MSO-definable subshift for which its forbidden language is not enumerable. This means in particular that there exists no recursive set $\mathcal{F}$ of patterns that defines this subshift, and in particular no *MSO-definable* set of patterns that defines this subshift. We will show in this section that this situation does not happen for the classes of subshifts we show before, that is every subshift of theses classes can be defined by a set of forbidden patterns of the same (logical) complexity.

For this to work, we now consider a purely relational signature, that is we consider now $\mathsf{East}, \mathsf{North}, \mathsf{South}, \mathsf{West}$ as binary relations rather that functions. As we said before, the previous theorems with the exception of theorem H.6 are still valid in this context. However with a relational signature, it makes sense to ask whether a given (finite) pattern P satisfy a formula ϕ: First-order quantifiers range over $\operatorname{Dom} P$, the domain of P, and second-order monadic quantifiers over all subsets of $\operatorname{Dom} P$.

We now prove

Theorem H.13

Let ϕ be a formula of the form

$$\exists/\forall X_1, \exists/\forall X_2 \ldots, \exists/\forall X_n, \forall z_1, \ldots, \forall z_p,\ \psi(X_1, \ldots, X_n, z_1 \ldots z_p)$$

Then a configuration M satisfies ϕ if and only if all patterns P of M satisfy ϕ.

Proof : The basic idea is to use compactness to bypass the existential (second-order) quantifiers.

We denote by $E_{\operatorname{Dom} P}$ the restriction of E to $\operatorname{Dom} P$. We prove the following statement by induction: For every subsets $E_1 \ldots E_k$ of $\mathbb{Z}^d$ and any configuration M, $(M, E_1, \ldots, E_k) \models \phi(X_1 \ldots X_k)$ if and only if $(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}) \models \phi(X_1 \ldots X_k)$ for every pattern P of M.

This is clear if ϕ has no second-order quantifiers.

Now let ϕ be a formula of the previous form. Note that it is clear that if $(M, E_1, \ldots, E_k) \models \phi(X_1 \ldots X_k)$ then $(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}) \models \phi(X_1 \ldots X_k)$, as the first order fragment of ϕ is universal. We now prove the converse. There are two cases:

- First case, $\phi(X_1 \ldots X_k) = \forall X \psi(X_1 \ldots X_k, X)$. Now suppose that

$$(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}) \models \phi(X_1 \ldots X_k)$$

for every pattern P of M. Let E be a subset of $\mathbb{Z}^d$. Now,

$$(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}, E_{\operatorname{Dom} P}) \models \psi(X_1 \ldots X_k, X)$$

for all patterns P of M by hypothesis, so using the induction hypothesis, we obtain $(M, E_1, \ldots, E_k, E) \models \psi(X_1 \ldots X_k, X)$, hence the result $(M, E_1 \ldots E_k) \models \forall X \phi(X_1 \ldots X_k, X)$.

- Second case, $\phi(X_1 \ldots X_k) = \exists X \psi(X_1 \ldots X_k, X)$. Suppose that for every pattern P of M, we have

$$(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}) \models \phi(X_1 \ldots X_k)$$

For every pattern P, we can find a set E_P so that $(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}, E_P)$ satisfies $\psi(X_1, \ldots X_k, X)$

Let P_i be the pattern of domain $[-i,i]^d$ of M, and $E_{P_i} \subseteq [-i,i]$ the subset given by the previous sentence. We now see E_{P_i} as a point in $\{0,1\}^{\mathbb{Z}^d}$, and by compactness we know that the set $\{E_{P_i}, i \in \mathbb{N}\}$ has an accumulation point E. This set E has the following property: for every domain $Z \subseteq \mathbb{Z}^d$, there exists i so that $[-i,i]^d$ contains Z, and E_{P_i} and E coincide on Z.

Now we prove that $(M, E_1, \ldots E_k, E)$ satisfies ψ. Let P be a pattern of M. There exists i so that E_{P_i} and E coincide on $\operatorname{Dom} P$. Now by definition of E_{P_i}, we have

$$(P_i, (E_1)_{\operatorname{Dom} P_i}, \ldots, (E_k)_{\operatorname{Dom} P_i}, E_{P_i}) \models \psi(X_1 \ldots X_k, X)$$

However, as P is a subpattern of P_i, and the fact that the first order fragment of ψ is universal, we have that

$$(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}, (E_{P_i})_{\operatorname{Dom} P}) \models \psi(X_1 \ldots X_k, X)$$

Now E coincide with E_{P_i} on $\operatorname{Dom} P$, so that we have

$$(P, (E_1)_{\operatorname{Dom} P}, \ldots, (E_k)_{\operatorname{Dom} P}, E_{\operatorname{Dom} P}) \models \psi(X_1 \ldots X_k, X)$$

We now obtain $(M, E_1, \ldots E_k, E) \models \psi(X_1 \ldots X_k, X)$ with the induction hypothesis, hence $(M, E_1 \ldots E_k \models \exists X \psi(X_1 \ldots X_k, X)$. ■

Corollary H.14

If S is a subshift defined by a formula ϕ of the form of the preceding theorem, then $S = X_{\mathcal{F}}$ where $\mathcal{F}$ is the set of words that do not satisfy ϕ.

In particular, in dimension 1, if a subshift is defined by a EMSO formula (is sofic), then it is defined by a EMSO-definable set of forbidden words, ie a regular set. Similarly, if a subshift is defined by a (universal) FO formula, it is defined by a (universal) FO-definable set of forbidden words, hence in particular by a strongly threshold locally testable language [BP91] (compare with corollary H.9).

The previous section shows that the corollary does not work for arbitrary formula ϕ. Indeed, for any MSO-formula ϕ, the set of words that do not satisfy ϕ is recursive, but there exists MSO-definable subshifts that cannot be given by a recursive set of forbidden words.

H.5.3 Definability of MSO-subshifts

As we saw before, sets defined by MSO-formulas are not always subshifts. We will try in this section to find a fragment of MSO that contains only subshifts and contain all of them. This fragment is somewhat ad hoc. Finding a more reasonable fragment is an interesting open question.

We first begin by a definition

Definition H.7

$$fin(S) : \exists A, \exists B \begin{cases} \forall x, A(x) \iff A(\mathsf{North}(x)) \wedge A(\mathsf{East}(x)) \\ \forall x, B(x) \iff A(\mathsf{South}(x)) \wedge A(\mathsf{West}(x)) \\ \exists x, A(x) \wedge \neg A(\mathsf{South}(x)) \wedge \neg A(\mathsf{West}(x)) \\ \exists x, B(x) \wedge \neg B(\mathsf{North}(x)) \wedge \neg B(\mathsf{East}(x)) \\ \forall x, S(x) \implies A(x) \wedge B(x) \end{cases}$$

It is easy to prove that $fin(S)$ is true if and only if S is finite (there are finitely many x such that $S(x)$). Indeed A and B represent quarter of planes, and S must be contained in the square delimited by the two quarter of planes. Any other formula true only if S is finite would work in the following.

Theorem H.15

Let X be a MSO-definable set. Then X is a subshift if and only if it is definable by a formula of the form

$$\forall S, fin(S) \implies \exists B_1 \dots B_k, \psi(S, B_1 \dots B_k) \wedge \forall x_1 \dots x_n S(x_1) \wedge \dots S(x_p) \implies \theta(S, B_1 \dots B_k, x_1 \dots x_p)$$

where
- ψ is any MSO-formula not containing the predicates P_c.
- θ is quantifier-free.

Note that this formula can be written more concisely as

$$\forall^{fin} S, \exists \overline{B} \psi(S, \overline{B}) \wedge \forall \overline{x} \in S^p, \theta(S, \overline{B}, \overline{x})$$

Proof : First we prove that such a formula ϕ defines a subshift X . For this, we prove that the set X is closed. Consider a sequence $M_1 \dots M_n \dots$ of configurations of X converging to some configuration M. We must prove that $M \in X$.

Let S be a finite set. Now consider the formula θ. As it is quantifier-free, it is local: the value of $\theta(S, B_1 \dots B_k, x_1 \dots x_n)$ depends only of what happens around $x_1 \dots x_n$. As each $x_1 \dots x_n$ must be in S, there exists a finite $S' \supset S$ such that the value of $\forall x_1 \in S \dots x_n \in S, \theta(S, B_1 \dots B_k, x_1 \dots x_n)$ depends only of the value of the predicates S, P_c and B_i on S'.

Now M_i converges to M. This means that there exists p such that M_p and M coincides on S'. For this M_p, there exists some $B_1 \dots B_k$ such that

$$\mathfrak{M}_p \models \psi(S, B_1 \dots B_k) \wedge \forall x_1 \in S \dots \forall x_p \in S, \theta(S, B_1 \dots B_k, x_1 \dots x_n)$$

Then this formula is also true on $\mathfrak{M}$ (Note indeed that $\psi(S, B_1 \dots B_k)$ does not depend on the configuration).

Hence we have found for every S some B_i that makes the formula true, that is we have proven $\mathfrak{M} \models \phi$. Therefore X is closed, hence a subshift.

Now let X be a MSO-definable subshift. X is defined by a formula ϕ. Change each P_c in ϕ by a predicate B_c to obtain ψ_1. Define

$$\psi(\overline{B}) = \forall x \left(\bigvee_c B_c(x) \right) \wedge \left(\bigwedge_{c \neq c'} \neg (B_c(x) \wedge B_{c'}(x)) \right) \wedge \psi_1(\overline{B})$$

Then X is defined by

$$\phi : \forall^{fin} S, \exists \overline{B} \psi(\overline{B}) \wedge \forall x \in S, \bigwedge_c (B_c(x) \iff P_c(x))$$

Indeed M satisfies ϕ and only if every pattern of M is a pattern in some configuration of X. ■

H.6 A Characterization of EMSO

EMSO-definable sets are projections of FO-definable sets (proposition H.1). Besides, sofic subshifts are projections of subshifts of finite type (or tilings). Previous results show that the correspondence sofic↔EMSO fails. However, we will show in this section how EMSO can be characterized through projections of "locally checkable" configurations.

Corollary H.4 expresses that FO-definable sets are essentially captured by counting occurrences of patterns up to some value. The key idea in the following is that this counting can be achieved by local checkings (equivalently, by tiling constraints), provided it is limited to a finite and explicitly delimited region. This idea was successfully used in [GRST96] in the context of picture languages: pictures are rectangular finite patterns with a border made explicit using a special state (which occurs all along the border and nowhere else). We will proceed here quite differently. Instead of putting special states on borders of some rectangular zone, we will simply require that two special subsets of states Q_0 and Q_1 are present in the configuration: we call a (Q_0, Q_1)-*marked configuration* any configuration that contains both a color $q \in Q_0$ and some color $q' \in Q_1$ somewhere. By extension, given a subshift Σ over Q and two subsets $Q_0 \subseteq Q$ and $Q_1 \subseteq Q$, the *doubly-marked set* Σ_{Q_0,Q_1} is the set of (Q_0, Q_1)-marked configurations of Σ. Finally, a *doubly-marked set of finite type* is a set Σ_{Q_0,Q_1} for some SFT Σ and some Q_0, Q_1.

Lemma H.16

Consider any finite pattern P and any $k \geq 0$. Then $S_{=k}(P)$ is the projection of some doubly-marked set of finite type. The same result holds for $S_{\geq k}(P)$.

Moreover, any positive combination (union and intersection) of projections of doubly-marked sets of finite type is also the projection of some doubly-marked sets of finite type.

Proof (sketch): For the first part of the theorem statement, we consider some base alphabet Q, some pattern P and some $k \geq 0$. We will build a doubly-marked set of finite type over alphabet $Q' = Q \times Q_+$ and then project back onto Q. The set Q_+ is itself a product of different layers. The first layer can take values $\{0, 1, 2\}$ and is devoted to the definition of the marker subsets Q_0 and Q_1: a state is in Q_i for $i \in \{0, 1\}$ if and only if its value on the first layer is i.

We first show how to convert the appearance in a configuration of two marked positions, by Q_0 and Q_1, into a locally identifiable rectangular zone. The zone is defined by two opposite corners corresponding to an occurrence of some state of Q_0 and Q_1 respectively. This can be done using only finite type constraints as follows. By adding a new layer of states, one can ensure that there is a unique occurrence of a state of Q_0 and maintain everywhere the following information:

1. $N_{Q_0}(z) \equiv$ the position z is at the north of the (unique) occurrence of a state from Q_0,
2. $E_{Q_0}(z) \equiv$ the position z is at the east of the occurrence of a state from Q_0.

The same can be done for Q_1. From that, the membership to the rectangular zone is defined at any position z by the following predicate (see figure H.6):

$$Z(z) \equiv N_{Q_0}(z) \neq N_{Q_1}(z) \wedge E_{Q_0}(z) \neq E_{Q_1}(z).$$

We can also define locally the border of the zone: precisely, cells not in the zone but adjacent to it. Now define $P(z)$ to be true if and only if z is the lower-left position in an occurrence of the pattern P. We add k new layers, each one storing (among other things) a predicate $C_i(z)$ verifying

$$C_i(z) \Rightarrow Z(z) \wedge P(z) \wedge \bigwedge_{j \neq i} \neg C_j(z).$$

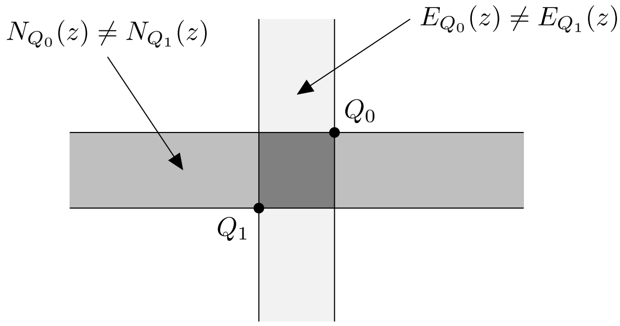


Figure H.4: The rectangular zone in dark gray defined by predicate $Z(z)$.

Moreover, on each layer i, we enforce that exactly 1 position z verifies $C_i(z)$: this can be done by maintaining north/south and east/west tags (as for Q_0 above) and requiring that the north (resp. south) border of the rectangular zone sees only the north (resp. south) tag and the same for east/west. Finally, we add the constraint:

$$P(z) \wedge Z(z) \Rightarrow \bigvee_i C_i$$

expressing that each occurrence of P in the zone mut be "marked" by some C_i. Hence, the only admissible (Q_0, Q_1)-marked configurations are those whose rectangular zone contains exactly k occurrences of pattern P. We thus obtain exactly $S_{\geq k}(P)$ after projection onto Q. To obtain $S_{=k}(P)$, it suffices to add the constraint:

$$P(z) \Rightarrow Z(z)$$

in order to forbid occurrences of P outside the rectangular zone.

To conclude the proof we show that finite unions or intersections of projections of doubly-marked sets of finite type are also projections of doubly-marked sets of finite type. Consider two SFT X over Q and Y over Q' and two pairs of marker subsets $Q_0, Q_1 \subseteq Q$ and $Q'_0, Q'_1 \subseteq Q'$. Let $\pi_X : Q \to A$ and $\pi_Y : Q' \to A$ be two projections. Denote by Σ_X and Σ_Y (resp.) the subsets of $A^{\mathbb{Z}^2}$ defined by $\pi_X\left(X_{Q_0,Q_1}\right)$ and $\pi_Y\left(Y_{Q'_0,Q'_1}\right)$. We want to show that both the union $\Sigma_x \cup \Sigma_Y$ and the intersection $\Sigma_X \cap \Sigma_Y$ are projections of some doubly marked sets of finite type.

First, for the case of union, we can suppose (up to renaming of states) that Q and Q' are disjoint and define the SFT Σ over alphabet $Q \cup Q'$ as follows:
- 2 adjacent positions must be both in Q or both in Q';
- any pattern forbidden in X or Y is forbidden in Σ.

Clearly, $\pi(\Sigma_{Q_0 \cup Q'_0, Q_1 \cup Q'_1}) = \pi_X(X_{Q_0,Q_1}) \cup \pi_Y(Y_{Q'_0,Q'_1})$ where $\pi(q)$ is $\pi_X(q)$ when $q \in Q$ and $\pi_Y(q)$ else.

Now, for intersections, consider the SFT Σ over the fiber product

$$Q_\times = \{(q, q') \in Q \times Q' | \pi_X(q) = \pi_Y(q')\}$$

and defined as follows: a pattern is forbidden if its projection on the component Q (resp. Q') is forbidden in X (resp. Y);

If we define π as π_X applied to the Q-component of states, and if E is the set of configuration of Σ such that states from Q_0 and Q_1 appear on the first component and states from Q'_0 and Q'_1 appear on the second one, then we have:

$$\pi(E) = \pi_X(X_{Q_0,Q_1}) \cup \pi_Y(Y_{Q'_0,Q'_1}).$$

To conclude the proof, it is sufficient to obtain E as the projection of some doubly-marked set of finite type. This can be done starting from Σ and adding a new component of states whose behaviour is to define a zone from two markers (as in the first part of this proof) and check that the zone contains occurrences of Q_0, Q_1, Q'_0 and Q'_1 in the appropriate components. ■

Theorem H.17

A set is EMSO-definable if and only if it is the projection of a doubly-marked set of finite type.

Proof : First, a doubly-marked set of finite type is an FO-definable set because SFT are FO-definable (theorem H.6) and the restriction to doubly-marked configurations can be expressed through a simple existential FO formula. Thus the projection of a doubly-marked set of finite type is EMSO-definable.

The opposite direction follows immediately from proposition H.1 and corollary H.4 and the lemma above. ■

At this point, one could wonder whether considering simply-marked set of finite type is sufficient to capture EMSO via projections. In fact the presence of 2 markers is necessary in the above theorem: considering the set Σ_{Q_0,Q_1} where Σ is the full shift $Q^{\mathbb{Z}^2}$ and Q_0 and Q_1 are distinct singleton subsets of Q, a simple compactness argument allows to show that it is not the projection of any simply-marked set of finite type.

H.7 Open Problems

- Is the second order alternation hierarchy strict for MSO (considering our model-theoretic equivalence)?
- One can prove that theorem H.6 also holds for formulas of the form:

  $$\forall X_1 \ldots \forall X_n, \forall z,\ \psi(z, X_1 \ldots X_n)$$

  where ψ is quantifier-free. Hence, adding universal second-order quantifiers does not increase the expression power of formulas of theorem H.6. More generally, let $\mathcal{C}$ be the class of formulas of the form

  $$\forall X_1, \exists X_2, \ldots, \forall/\exists X_n, \forall z_1, \ldots, \forall z_p, \phi(X_1, \ldots, X_n, z_1, \ldots, z_p).$$

  One can check that any formula in $\mathcal{C}$ defines a subshift. Is the second-order quantifiers alternation hierarchy strict in $\mathcal{C}$? On the contrary, do all formulas in $\mathcal{C}$ represent sofic subshifts ?